IDA

INSTITUTE FOR DEFENSE ANALYSES

Integrated Cognition – A Proposed Definition of Ingredients, A Survey of Systems, and Example Architecture

Robert M. Rolfe, Task Leader

Brian A. Haugh

October 2004

Approved for public release;
unlimited distribution.

IDA Paper P-3855

Log: H 03-002468

This work was conducted under contracts DASW01 98 C 0067/ DASW01 02 C 0012/W74V8H 05 C 0042, Task DA-5-2326, for the Defense Advanced Research Projects Agency (DARPA). The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-3855

Integrated Cognition – A Proposed Definition of Ingredients, A Survey of Systems, and Example Architecture

Robert M. Rolfe, Task Leader

Brian A. Haugh

Preface

This document was prepared by the Institute for Defense Analyses (IDA) under the task order Integrated Cognition for the Defense Advanced Research Projects Agency (DARPA). It was developed in response to task objectives to: 1) develop a multi-dimensional framework for Integrated Cognition (INCOG), 2) perform a qualitative technical assessment of selected existing cognitive architectures with respect to this framework, and 3) to develop a strawman example of a cognitive architecture that supports the capabilities of this framework.

We thank Dr. Ronald Brachman and Dr. Barbara Yoon of the DARPA Information Processing Technology Office (IPTO) for their support of this work. We thank Dr. John Salasin for his vision in conceiving these investigations into INCOG. We would also like to thank Professor Pat Langley of the Computational Learning Laboratory at Stanford University, who motivated Dr. Salasin and Dr. Rolfe to develop a key diagram to communicate the scope and difficulty of the undertaking to construct intelligent machine architecture. We also extend our gratitude to the many researchers in cognitive architectures who assisted in assessing their architectural coverage of the ingredients of cognition identified in the INCOG framework presented herein, including: Dr. John R. Anderson, Mr. Albert-Laszlo Barabasi, Dr. Michael E. Bratman, Dr. Gerald M. Edelman, Professor Kenneth D. Forbus, Dr. Chris Furmanski, Professor Dedre Gentner, Dr. Michael P. Georgeff, Dr. Ben Goertzel, Professor Marvin Minsky, Dr. Robert Hecht-Nielsen, Dr. Marcus J. Huber, Dr. John Laird, Professor Pat Langley, Dr. Christian Lebiere, Dr. David J. Musliner, Ms. Karen L. Myers, Professor Matthias Scheutz, Mr. Push Singh, Dr. Aaron Sloman, Mr. Lokendra Shastri, Dr. William R. Swartout, Dr. Manuela Veloso, Dr. Robert Wray, and Dr. H. Van Dyke Parunak.

The following IDA research staff members were reviewers of this document: Dr. Donald J. Goldstein, Mr. Louis Lome, Dr. L. Roger Mason, Jr., and Mr. Michael G. Sullivan. In addition, Mr. Push Singh of the MIT Media Lab served as an external reviewer.

Contents

List of Figures

List of Tables

Executive Summary

Problem

The many substantial benefits of human-level machine intelligence for national security applications alone argues for greater investment in cognitive architectures for truly intelligent systems. While considerable progress has been made in many areas of cognitive science and intelligent systems research, we do not yet have computer systems with the intelligence of a five year old child.

One problem faced by DARPA, as the principal U.S. DoD agency for advanced research, is how best to foster development of human-level intelligence in computer systems to enable realization of its many potential benefits for our national security.

Approach

The approach proposed here is based on a framework for "integrated cognition" (INCOG) which can be used both to organize essential capabilities of human-level cognition, and to assess existing and proposed architectures and systems on their progress in realizing these capabilities. This INCOG framework is used in this paper to assesse a score of existing cognitive architectures.

The proposed INCOG approach is to specify architectural strategies and develop infrastructure mechanisms (if needed) that enable artificial systems to address all the cognitive requirements of the INCOG framework. These architectural strategies should provide languages and tools for expressing the interaction of cognitive components. They should make it easier to compose the components to solve a specific problem or for the system itself to understand how it operates so it can compose itself as needed.

As an illustration of this INCOG strategy, a strawman INCOG architecture is presented which seeks to integrate all facets of all the identified dimensions of cognition in the INCOG framework.

INCOG Framework

The INCOG framework is based on the premise, shared by numerous cognitive and other scientists, that a human-level thinking machine must be composed of many, potentially hundreds, of distinct components with different structures, reasoning and learning mechanisms, and knowledge representations. These components and their inter-relationships define a dynamic architecture or family of architectures.

Many researchers believe that a cognitive system should have an ability to learn using different approaches; an ability to reason in many domains; an ability to acquire and build knowledge; and an ability to connect, combine, integrate, collaborate and unify knowledge across many domains. These distinct, though related, abilities form the dimensions of the INCOG framework presented here.

Figure ES-1 presents the INCOG framework as a diagram with six axes or dimensions, each representing natural groupings of key ingredients (or capabilities) of integrated cognition. Each of the capabilities illustrated along these axes become (arguably) increasingly more difficult to obtain moving outward from the center to the periphery. Each dimension represents a challenge by itself; taken together, they present a "DARPA hard problem."

The INCOG framework described here is presented only as one possible way of parsing cognitive capabilities into a coherent framework. Many alternative frameworks are possible. While the current framework has benefited from the inputs of a number of researchers in cognitive architectures, it is not presented as a definitive model but only as a starting point. But something like this is needed to provide a basis for assessing existing cognitive architectures, a guide to research investment, and a measure of progress towards the grail of truly intelligent systems.

INCOG Framework Dimensions

The INCOG dimension of **Multi-level Mind** is based on Professor Marvin Minsky's distinctions of a six level model of mind, as reported in a recent draft of his forthcoming book "The Emotion Machine" [MINSKY 2004]. The multi-level mind model deals with a number of issues, from the relationship of response time and cognitive processes to possible levels of contending control and management in the mind.:

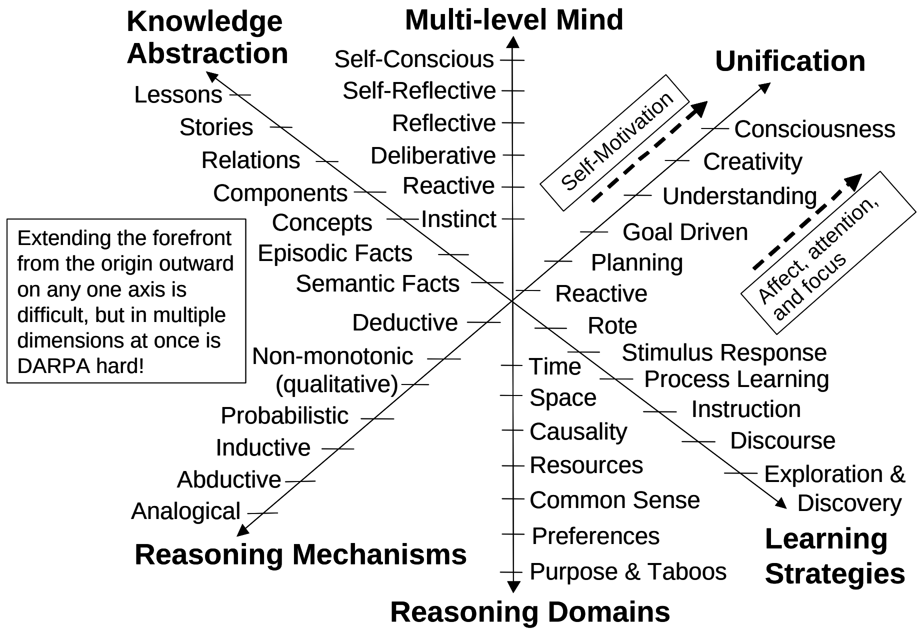

Figure ES-1. INCOG Framework for Ingredients of Integrated Cognition

Unification is the process of combining the different elements in a cognitive system to generate a cognitive agent's behavior. Different types of unification are distinguished in part by the level of the "multi-level mind" at which they operate. But other aspects of an instantiated cognitive architecture, such as purpose and taboos, will affect the focus of unification. Any unification level may draw on lower levels of unification as needed for various behaviors.

Learning Strategies of artificial cognitive systems are still quite primitive compared to human learning. However, fully capable integrated cognition systems should be able to utilize all of the learning strategies included in the INCOG framework. Different strategies for learning are distinguished here largely by the amount and types of inference required of the learner during the course of learning.

The INCOG framework provides an ordered list of general purpose **Reasoning Domains**. Included in the challenges of each reasoning domain is the representation of information, as well as the formulation and implementation of special purpose reasoning algorithms for the domain. It can be argued that these challenges increase as we move outwards on the Reasoning Domains axis. Spatial reasoning, for example, can be considered as more complex than temporal reasoning since it commonly involves up to three dimensions, whereas time ordinarily has but one.

Reasoning Mechanisms are systems for generating conclusions based on premises. This dimension includes specification of information and its associated semantics, inference rules, and algorithms for generating conclusions. Probabilistic reasoning, for example, requires specification with numerical qualifiers to indicate the probabilities of premises and conclusions.

Knowledge Abstraction provides a classification of knowledge in levels of increasing complexity, from simple definitional facts to complex stories and lessons. Knowledge is the collection of persisting memories used by cognitive systems to support reasoning, learning, and unification about the internal and external world.

A Survey of Systems

This paper presents individual assessments of the coverage of INCOG framework capabilities by over fifteen prominent cognitive architectures or systems. These assessments were performed in collaboration with the system architects, as cited below. Each system was assessed for its support of each capability in the INCOG framework. These assessments, individually and collectively, provide a view of the current state of cognitive architectures relative to one model of the requirements of human-level cognition.

In order to keep this summary brief, it does not include the individual system assessments, although brief descriptions of the examined

systems are provided to give some idea of the scope.

Soar is a general purpose architecture designed as an unified theory of cognition by John Laird, Paul Rosenbloom, and Allen Newell. It is a production rule system based on the simple decision cycle - elaboration of state, choice of operators, selection of operator, and actions. Soar has a relatively large user base among existing cognitive architectures. It is supported by the University of Michigan and has been applied commercially by Soar Technology Inc.

ACT-R is a cognitive architecture using production rules developed at Carnegie Mellon University by John Anderson and Christian Lebiere. It includes a detailed approach to integrating multiple modules that correspond to different cognitive functions. The fundamental controlling structure in cognition is reactive– where production rules respond to patterns of information in various cognitive buffers. Successive versions of ACT-R have seen widespread applications to problems of cognitive and behavioral modeling.

The distributed Multi-Agent Reasoning System (**dMARS**) is a C++ implementation of an architecture based on the Belief, Desire, Intention (BDI) cognitive model [IKLW 1998]. It was developed by Michael Georgeff as a more powerful successor to the Procedural Reasoning System (PRS). dMARS has been applied to a very wide range of applications, including command and control of robotics and spacecraft; and situation awareness for the Australian Defense Forces.

ICARUS is an architecture for intelligent agents developed by Dan Shapiro and Pat Langley of the Center for the Study of Language and Information at Stanford University. ICARUS is distinguished by its incorporation of affective values into memory and behavior; the primacy of categorization over execution and of execution over problem solving; and the internal determination of tasks, intentions, and rewards.

DARWIN refers to a series of implementations of large-scale (over 50,000 cells and 600,000 synapses) synthetic models of neural structures supporting the evolution of pattern recognition and sensorimotor coordination in a synthetic environment. It has been developed by Reeke, Sporns, and Edelman of the Neurosciences Institute and Rockefeller University based on Edelman's theories of Neural Darwinism.

UMPRS (the University of Michigan implementation of PRS) is a general purpose implementation of PRS. It does not provide (i.e., "impose") specific capabilities or representations on agent programmers, but rather provides a framework for their implementation. Hence, its core capabilities cover relatively few of the INCOG framework ingredients, although UMPRS applications have covered many more. Unification in UMPRS is focused on goals and planning and not reactive tasks.

Shruti and Smirti are related architectures developed by Lokendra Shastri of UC Berkeley [SHAST 1999]. They demonstrate how simple, neuron-like, elements can encode a large body of relational causal knowledge and provide a basis for reactive, rapid inference.

SAGE (Self-Aware Adaptive Generalization Engine) is a cognitive architecture developed by Chris Furmanski and John Hummel of UCLA. It is adaptive, self-reliant, and can reason by analogy in order to discover meaningful relationships between seemingly dissimilar data. It blends connectionist/neural networks with symbolic systems. Its self-supervised learning uses self-reflective algorithms that allow the system to acquire new knowledge and learn from its past.

"Panalogy is a cognitive architecture designed to support commonsense reasoning being developed by Push Singh and Marvin Minsky of MIT." The Panalogy system makes it possible to connect multiple representations of knowledge and combine diverse reasoning techniques such as analogy, case-based reasoning, statistical estimation, and logical inference. It does so by applying a wide array of meta-managerial components that use meta-knowledge about how to select, coordinate and repair baseline reasoning and learning processes.[1]"

Novamente is a system organized with distributed atoms of knowledge that may be employed in an unlimited number of contexts developed by Ben Goertzel of Artificial General Intelligence Research Institute (AGIRI). Atoms have truth value and attention value. Mind agents operate on these atoms, learning how to learn.

JAM is another version of the PRS, this one developed by Marcus Huber of Intelligent Reasoning Systems. It provides Reactive and Deliberative models as well as Reflective and Self-Reflective capabilities in the form goal semantics and meta-level reasoning. Its unification is limited to planning and goal-driven behavior.

Daydreamer was developed by Erik Mueller of IBM to simulate a human stream of thought and its triggering and direction by emotions.

The tools **SME, SEQL** and **MAC/FAC** have been developed at Northwestern University by

[1] Private communication, Push Singh, MIT Media Lab.

Ken Forbus, Dedre Gentner, and others. These have been designed for structure-mapping and qualitative reasoning.

ThoughtTreasure is a story understanding and commonsense reasoning system developed by Erik Mueller of IBM.

Stigmergic Cognition consists of cognitive components that exhibit emergent behavior and have performed well in many roles. It has been developed by H. Van Parunak and Sven Brueckner of Altarum.

Survey Conclusions

General observations supported by this survey can be summarized as follows:

- Working implementations tend to require significant low-level programming for ingredients not already present in the architectural models (e.g., SOAR, ACT-R, dMARS, UMPRS)
- Learning capabilities are bounded to refinement within the scope of initial knowledge bases; few systems seek to understand, or invoke learning strategies beyond process learning related to current knowledge
- Current systems tend to be weak with respect to self-reflection and knowledge sharing and consequently would be difficult to employ in a heterogeneous integrated cognitive architecture without extensions
- Current systems have core capabilities that cluster near the center of the INCOG framework diagram
- Newer or proposed systems have extended coverage of the strawman multi-dimensional framework near the periphery in some dimensions
- There are significant sets of advanced components (from established cognitive systems) that could be included in integrated architectures that together would provide significant new capability for cognitive systems.

Strawman INCOG Architecture

The final part of this paper develops an example top-level architecture that meets the requirements of the strawman integrated cognition framework described herein. A top-level view of the INCOG strawman architecture is shown in Figure ES-2. This figure is an adaptation of Ronald Brachman's proposed cognitive architecture, designed to better capture distinctions made in the INCOG framework. In particular, this INCOG architecture distinguishes more finely between different levels of the multi-level mind, separating Brachman's *Reactive Processes* into *Programmed Instinct* and *Learned Reactions*; and adding several other levels above the *Reflective*. This architecture also highlights discourse as a key area of human-level cognition by separately identifying it and its relations to other input and output processing.

The world external to the cognitive agent is represented in the diagram by the *External World* box at the bottom. The rest of the boxes represent information and processes of the cognitive agent. Raw *Sensory Inputs* come into the agent from its sensors and are processed initially via *Perception Processing*, which hands off linguistic data to *Discourse Input*. The results of perception and discourse input processing are fed to various levels of the *Mult-ilevel Mind*, as appropriate. The *Multi-level Mind* uses inputs from *Working Memory* and *Long Term Memory* to place new inputs in context and to help determine its responses and other activities. The *Multi-level Mind* also stores information in *Working Memory* and *Long-Term Memory* as warranted. The processing and exchange of information throughout is enabled by a host of *Foundation Processes*, as well as the *Computational Substrate*. The results of the execution of cognitive processes may then find expression in discourse via *Discourse Output* or through other *Action Outputs*.

To identify more of the specific functional capabilities required to instantiate this architecture, a number of functional components were conceived and grouped into categories correlated with the INCOG framework and architecture. The package groupings described in the paper are the following:

- Multi-Level Mind Packages
- Foundation Multi-Level Mechanisms
- Foundation Unification Packages
- Foundation Global Mechanisms Packages
- Foundation Reasoners Packages
- Foundation Domain Packages
- Foundation Discourse Support Packages

Two sets of packages are distinguished for the multi-level mind. The first applies only to the multi-level mind, while the second grouping contains foundation packages that may also be used by other parts of the INCOG architecture. The rest of these categories contain foundational packages that are represented by the *Foundational Processes* box on the top-level architecture diagram.

The paper describes a total of 61 different functional packages under these groupings. It then proceeds to explain how they can be composed to support different higher-level cognitive processes.

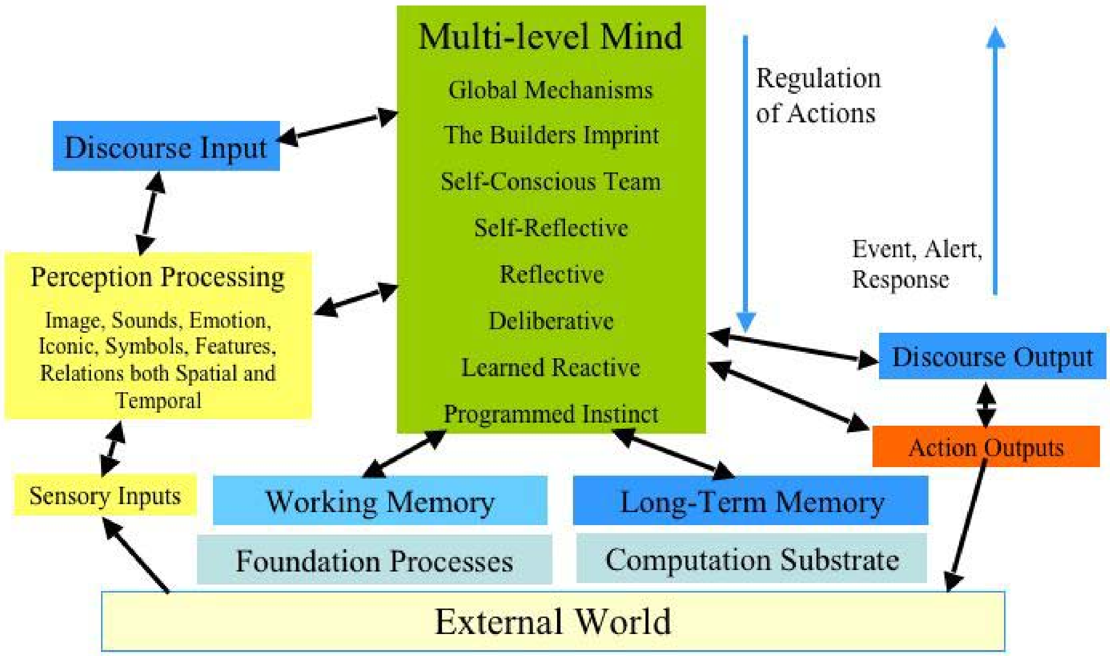

Figure ES-2. INCOG Strawman Architecture Top-Level Static Components

1. Introduction

1.1. Background

There is an expectation that computer-based systems will eventually exhibit or surpass the intelligent behavior of humans. There has been much hope in the cognitive science community that there is a unifying strategy that will enable us to achieve this vision. Numerous cognitive scientists are coming to the conclusion that a human-level thinking machine must be composed of potentially hundreds of distinct subsystems with different structures (architectures), reasoning and learning mechanisms, and knowledge representations.

A research program on Integrated Cognition (INCOG) has been conceived by the Defense Advanced Research Projects Agency (DARPA) that would develop architectural strategies[2] that enable computation based systems to achieve a reasonable facsimile of human cognition. The target system should demonstrate human-like capabilities with respect to selecting and employing appropriate mechanisms for learning and reasoning in different contexts. It would initially provide demonstrations of learning/reasoning capabilities at the level of a 5-year-old child, but lacking in experience. In years 3 and 4, it would demonstrate an ability to address deeper and more complex problems in areas relevant to military applications.

Many researchers believe that a cognitive system should have the following capabilities: the ability to learn using different approaches; the ability to reason in many domains; the ability to acquire and build knowledge; and the ability to connect, combine, integrate, collaborate, and unify knowledge across many domains.

1.2. Problem Statement

Many aspects of military operations in support of national security could benefit from human-level intelligence in automated systems. Intelligent information gathering, filtering, fusing, and mining by automated systems could contribute greatly to information superiority in the battlefield. The Future Combat System (FCS) and Objective Force will require highly autonomous behavior from intelligent robotic platforms. The more intelligence incorporated into these platforms and their information networks, the more effective such future forces will be. Acquisition, analysis, and training could all benefit from human-level intelligence in realistic computer models of forces used in their sup-

[2] Architecture includes principal components, interfaces, and principles of interrelationship.

porting modeling and simulation systems. Although the potential benefits of human-level computer intelligence for these and other DoD applications are obvious, we are still a long way from fully realizing such capabilities. Despite decades of research on knowledge representations, inference algorithms, and other technologies for intelligent systems, we do not yet have computer systems with the intelligence of a 5-year-old child.

One problem faced by DARPA, as the principal DoD agency for advanced research, is how best to foster development of human-level intelligence in computer systems to enable realization of its many potential benefits for the military. The related problem being addressed by work at IDA, reported in this paper, is how to provide technology analysis support to DARPA in addressing this problem of developing human-level machine intelligence.

1.3. Purpose

This paper reports on initial analyses of cognitive architectures that show some promise of contributing to the development of human-level intelligence in machines. It begins to explore alternative technical approaches for cognitive architectures inspired by psychological, biological, neurological, and other architectural concepts.

1.4. Approach

The approach to this initial analysis task involved the following basic steps:

1. Develop a strawman framework and paper defining the multiple dimensions of integrated cognition

2. Engage the research community to describe their cognitive systems (architectures) and components in terms of the strawman integrated framework

3. Develop a summary chart to illustrate potential capabilities for integrated cognition and iterate with each research team to refine as needed

4. Summarize hypothetical uses of surveyed systems in integrated cognition architectures

5. Sketch out a notional architecture that integrates all facets of all the identified dimensions of cognition.

The INCOG framework is described in Section 2, while the cognitive systems and architectures surveyed are reviewed in Section 3. The detailed survey results in terms of the framework's dimensions of cognition are provided in the charts of Appendix A. An introduction to composing a notional INCOG architecture is presented in Section 4, followed by the details in Appendix B.

2. Integrated Cognition Framework

The INCOG framework[3] for integrated cognition is based on the premise, shared by numerous cognitive and other scientists, that a human-level thinking machine must be composed of potentially hundreds of distinct subsystems with different structures, reasoning and learning mechanisms, and knowledge representations – with these components and their inter-relationships defining a (or a family of) dynamic architecture(s). The INCOG proposed approach is to specify architectural strategies and develop infrastructure mechanisms (if needed) that enable computation-based systems to achieve a reasonable facsimile of human cognition. The architectural strategies should provide languages and tools for expressing the interaction of cognitive components, ranging in size and complexity from large, tightly integrated systems to simpler agents in Multi-Agent Systems. Architecture strategies should make it easier to compose the components to solve a specific problem or for the system itself to understand how it operates so it can compose itself as needed.

Figure 1 illustrates some of the ingredients of integrated cognition needed to approach human like behavior.[4] Each of the axes represents a challenge by itself; taken together, they present a DARPA hard problem. One approach to future cognitive systems is to assure that whatever is learned to solve a problem today can be re-applied (where appropriate) to solve future problems. This might be achieved in part by building cognitive systems that reuse the variety of data and procedural abstractions needed to cover the large space illustrated in Figure 1. Each of the ingredients illustrated along the many axes in Figure 1 become increasingly difficult to obtain moving outward from the center to the edges. Further, the ability to re-apply ingredients requires as much independence as possible between the ingredients illustrated on the various axes.

Many researchers believe that a cognitive system should have the ability to learn using different approaches, the ability to reason in many domains, the ability to acquire and build knowledge, and the ability to connect, combine, integrate, collaborate, and unify knowledge across many domains. We will briefly discuss these areas in turn.

---

[3] In support of the Information Processing Technology Office at DARPA, IDA's Information Technology and Systems Division has drafted a framework for considering the scope of capabilities that must be integrated together to reach the INCOG goal.

[4] We would like to thank Professor Pat Langley of the Computational Learning Laboratory at Stanford University who motivated Dr. Salasin and Dr. Rolfe to develop Figure 1 to communicate the scope and difficulty of the undertaking to construct an intelligent machine architecture.

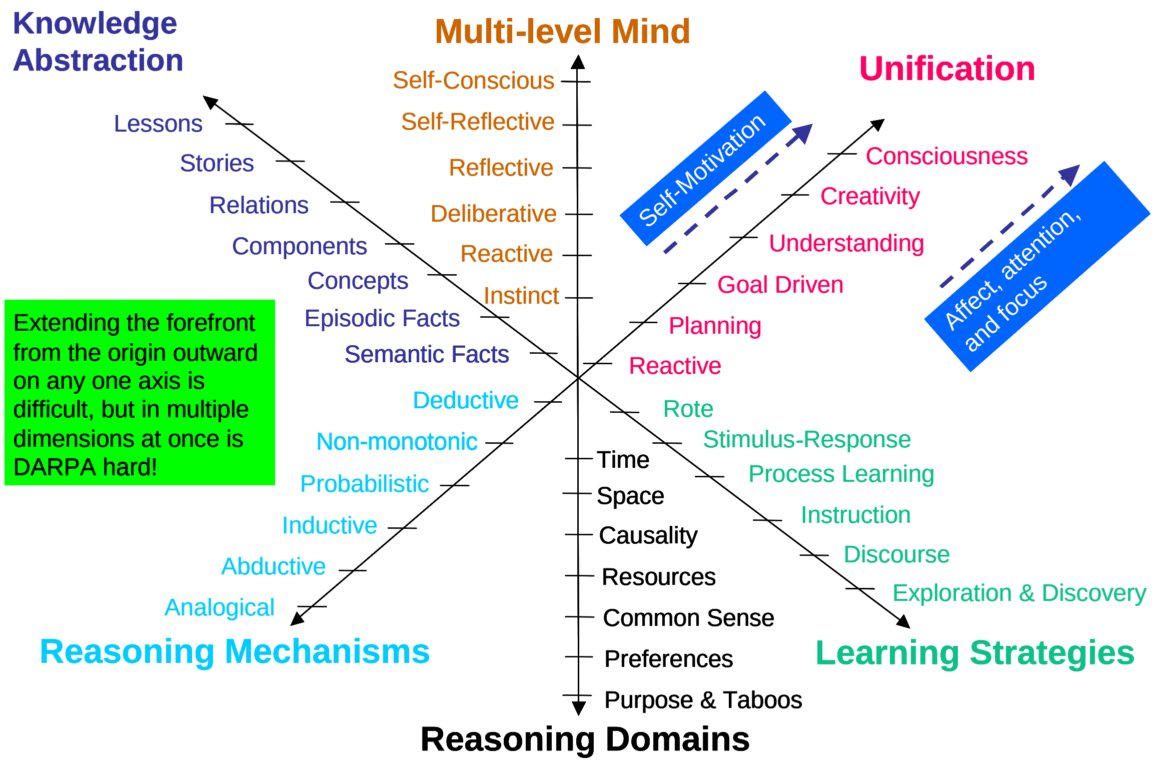

Figure 1. Framework for ingredients of integrated cognition

The sub-sections that follow provide a preliminary definition and or discussion of each ingredient in the context of the entire framework, as well as the dimension in which it is placed.

2.1. Multi-Level Mind

The quotations in this section are drawn from Chapter 5 of the January 2004 draft version of "*The Emotion Machine*", by Marvin Minsky [MNSKY 2004]. The multi-level mind model deals with a number of issues, from the relationship of response time and cognitive processes to possible levels of contending control and management in the mind. Minsky uses the following example of the thought processes of a woman crossing the street to illustrate how different levels of mind come into play in ordinary activities:

> *Joan is part way across the street on the way to present her finished report. While thinking about what to say at the meeting, she hears a sound and turns her head —and sees a quickly oncoming car. Uncertain whether to cross or retreat but uneasy about arriving late, Joan decides to sprint across the road. She later remembers her injured knee and reflects upon her impulsive decision. "If my knee had failed, I could have been killed—and what would my friends have thought of me?" [MNSKY 2004]*

2.1.1. Instinct

The level of Instinct (or Instinctive Reactions) is a level of mind based on reactions which are said to be "inborn" or genetic in humans and other animals. In machines,

operation at the level of Instinct corresponds to acting in accord with pre-programmed reactions, without any real deliberation or explicit planning. In the example, Joan*:*

> *hears a sound and turns her head.* Many infant animals do such things; they're born with just enough "instincts" to help them survive.

Machines may have instinctive responses programmed in terms of rules whose antecedents match against current conditions to provide a standard response, without any reasoning or inference about the consequences. Other implementation paradigms, such as neural nets or decision trees can also provide the basis for such pattern matching and instinctive reactions. Such reactive behavior can be considered "instinctive" in machines insofar as it is not learned through prior experience of the machine.

2.1.2. Reactive

The Reactive (or Learned Reactions) level in this framework is reserved for non-deliberative behavior that is learned by a cognitive agent. In the example, Joan:

> *sees a quickly oncoming car.* Joan has learned that certain conditions demand additional, specific reactions.

Operation at a Reactive level is characterized by quick responses (reactions) to sensory or other information that may be processed to varying degrees into representations of world situations. The Reactive level does not exclude formation and use of world or environmental models to guide behavior, so long as their use does not become deliberative (as described next). Joan's reactive response in our example may be simply her learned pause in response to recognition of speeding cars in her vicinity.

2.1.3. Deliberative

Minsky defines the Deliberative Thinking level in terms of our ordinary language understanding of the term. He acknowledges that "deliberation" is not precisely defined, but is rather a "suitcase" type term used to capture a variety of cognitive activities, such as predicting, comparing, and planning. Deliberation's consideration of alternative possible futures is one process that distinguishes it from the Reactive level. In the example, Joan is:

> *thinking about what to say at the meeting* Here she imagines alternative futures, and various ways to choose among them.

Distinguishing *Deliberative* from *Reactive* levels is important because deliberative consideration of alternative futures enables cognitive agents to solve more difficult problems than merely instinctive or reactive agents.

2.1.4. Reflective

Functioning at the Reflective level entails an agent representing and reasoning about what it has done and thought. The mind forms representations of its own action, thoughts, and deliberations, as well as the external world. In the example of Joan's thoughts:

> *She later remembers her injured knee and reflects upon her impulsive decision. . . .* now that she has more time to think, she can contemplate what she's recently done. Whenever you face a difficult problem, you may find yourself reflecting on your recent ideas, decisions, and actions.

2.1.5. Self-Reflective

The *Self-Reflective* level extends the *Reflective* level with some model of the entity (the *self*) that performed the deliberations and actions that it reflects upon. A cognitive agent operating at this level must have some model of itself, e.g., its capabilities and goals, and/or its place within various social contexts. Joan thinks:

> *"If my knee had failed, I could have been killed."* Here, Joan reflects on what she's been doing—and concludes that she made a poor decision: she should not have neglected her injured knee.

Self-reflective reasoning can improve an agent's functioning by enabling it to step back from a difficult problem or goal and reflect upon how it relates to other aspects of itself. Current goals are often derivative of broader interests that may be better served by changing or revising goals when they are frustrated by circumstances.

2.1.6. Self-Conscious

The Self-Conscious level extends self-reflection to reflections on how others perceive the self. It represents a capacity to consider and respond to the approval and disapproval of other individuals and social groups, using cultural values, goals, and taboos. As such, it is essential to full participation in societies of cognitive agents and to culture formation. In the example, Joan reflects:

> *"What would my friends have thought of me?"* Here Joan thinks both about something she's done—and whether she ought to have done it . . . To think such thoughts, Joan would have to possess resources that not only represent her body, as well as her various values and goals, but also her social relationships, . . .

The Self-Conscious level operates on all representations of the self and the world, and thus will result in behavior that includes cognition, affect, and motivation.

2.2. Unification

Unification is the process of combining the different elements in a cognitive system to generate a cognitive agent's behavior. Different types of unification are distinguished in part by the level of the "multi-level mind" at which they operate. But, other aspects of an instantiated integrated cognitive architecture, such as purpose and taboos, will affect the

focus of unification. It is assumed that any system unification level may draw on lower level unification ingredients as needed for various behaviors.

2.2.1. Reactive

Reactive unification involves nearly direct reaction to external stimuli. Systems that have only a reactive unification process would be difficult to consider as truly *cognitive* systems. A simple reactive system may be implemented by a set of reaction-stimuli pairs. However, not all reactive systems need be considered primitive since there may be very complex reasoning involved in their reactions. Further, reactive reasoning may occur in a number of domains and may use a knowledge abstraction and learning strategies. Reactive systems certainly would utilize the *Instinct* and/or *Reactive* elements of the Multi-Level Mind.

2.2.2. Planning

Planning-based unification provides a cognitive system behavior that incorporates planning elements. The behavioral response draws upon a set of task elements that can be grouped together to form alternative responses to external stimuli or current state information. Behavior might be characterized as extending over time while the system executes the plan. Behavior may driven by a set of measurable objectives. Systems with planning-based unification would certainly utilize some capabilities from the *Deliberative* level of the Multi-Level Mind. And, they may use the *Instinct* and/or *Reactive* levels of the mind as well. Planning-based systems might draw upon all other ingredient axes in the Figure 1 chart.

2.2.3. Goal-Driven

Goal-driven cognitive systems incorporate an understanding of goals that may extend over large periods of time and involve the generation of metrics to measure progress toward the goals. Such system may vary lower level unification processes to achieving a set of potentially conflicting goals. Still, such systems may not require broad-based coverage of knowledge, reasoning mechanisms, reasoning domains, and learning strategies. Goal-driven systems could utilize the *Instinct*, *Reactive*, *Deliberative*, and limited *Reflective* levels of the Multi-Level Mind.

2.2.4. Understanding

Cognitive systems with understanding unification attempt to increase understanding of the world and themselves. By understanding the nature of their own capabilities they may seek to perform behavior consistent with internal and external goals. Understanding processes may increase relevant knowledge or relate at-hand knowledge to new situations. Understanding as a unification mechanism would be expected to extend over the lifetime of an individual cognitive system. Further, such systems would be able to share abstract knowledge among teams of cognitive systems. Understanding cognitive systems will require broad-based coverage of knowledge, reasoning mechanisms, reasoning domains,

and learning strategies. Understanding unified cognitive systems would utilize the *Instinct, Reactive, Deliberative, Reflective* and *Self-Reflective* levels of the Multi-Level Mind.

2.2.5. Creativity

Cognitive systems with Creativity unification seek to not only perform behavior consistent with goals, but to understand the world about them and seek to extend knowledge and understanding about the unknown. Creativity-based unification of cognitive systems will result in systems that will understand, look for patterns and analogies, and explore the unknown. Such systems can speculate (form and test hypotheses) about the data and data relationships. They will require broad-based coverage of knowledge, reasoning mechanisms, reasoning domains, and learning strategies including discovery and exploration. Creativity unified cognitive systems would utilize the *Instinct, Reactive, Deliberative, Reflective, Self-reflective,* and limited *Self-conscious* levels of the Multi-Level Mind. Creativity requires synthesis of models and/or simulations that provide new insights; these insights may well involve advanced learning mechanisms such as "Exploration and Discovery" that would manipulate the environment to confirm or extend new mental models.

2.2.6. Consciousness

A cognitive system with consciousness unification is one that exhibits behavior that reflects the integrated cognition (thinking), affect, and motivation that so well characterizes human beings.[5] Behavior that is characteristic of consciousness is defined by Ortony et al. [ONR, In Prep] in terms of the traditional psychological triumvirate of Affection (affect), Conation (will, or motivation), and Cognition:[6]

> *Affect, motivation, and cognition can be considered to be the internal control mechanisms of the organism's behavior. We postulate that differences in steady state resting potentials, biases, thresholds, and weights for these mechanisms comprise the strikingly consistent and strikingly different organizations that are known as personality.*

These factors are indicated on the INCOG framework diagram (Figure 1) by a label conveying an increase of their involvement with increasing levels of unification. A cognitive system with such consciousness unification might well pass the Turing Test[7] with many

---

[5] Wording motivated by a private communication with Andrew Ortony, Donald A. Norman, and William Revelle based upon a draft of [ONR, in prep].

[6] Ibid, page 3.

[7] When talking about the Turing Test today what is generally understood is the following: The interrogator is connected to one person and one machine via a terminal, therefore she can't see her counterparts. The interrogator's task is to find out which of the two candidates is the machine, and which is the human only by asking them questions. If the machine can "fool" the interrogator, it is intelligent. http://cogsci.ucsd.edu/~asaygin/tt/ttest.html#intro

different individuals. It will most likely utilize all of the ingredients of cognitive systems illustrated in Figure 1. In particular it will require processes to support all levels of the Multi-Level Mind including the self-conscious.

2.3. Learning Strategies

Learning strategies of artificial cognitive systems are still quite primitive compared to human learning. However, fully capable integrated cognition systems should be able to utilize all of the learning strategies[8], [9] illustrated above in Figure 1.

The American Association for Artificial Intelligence (AAAI) website on machine learning states that machine learning is said to occur in a program that can modify some aspect of itself, often referred to as its state, so that on a subsequent execution with the same input, a different (hopefully better) output is produced [AAAI 2004]. Different strategies for learning may be distinguished by the amount and types of inference required of the learner during the course of learning.[10] Simply adding data to systems information stores can increase its knowledge, but this would be considered a very simple learning strategy. In contrast, a system capable of performing experiments and generalizing the results to new scientific theories exhibits a very sophisticated learning strategy. Higher forms of learning may depend upon the appropriate level of unification to be included in the cognitive system, e.g., creativity.

2.3.1. Rote Learning

Rote learning is the simplest type of learning strategy in which "no inference or other transformation of the knowledge is required on the part of the learner" [MCM 1983]. In artificial systems, rote learning can be accomplished by entry of data or knowledge by an external agent. In humans, rote learning typically involves memorization of facts without much depth of understanding or integration of them with other knowledge.

---

8 An overview of such machine learning strategies can be found in [MCM 1983], while a more in depth review is provided in [MITCHL 1997].

9 Private communication, Push Singh of MIT Media lab on machine learning, *"a powerful learning machine depends on having a powerful architecture—for (a) much of learning is about making improvements to your own architecture, and if your architecture does not allow sufficient variation in types of components and ways that those components can be interconnected, then only limited self-improvement is possible, and (b) to learn quickly requires great intelligence, in the sense that you need to be able to generate good hypotheses about the likely effects of changes to yourself and also good ways to choose between those hypotheses, and there may be no "general mechanism" for this besides applying the full intelligence of the resulting architecture to the problem (as opposed to limiting yourself to a few weak "learning heuristics" as all learning systems today do).*

10 This means of distinguishing learning strategies is adapted from [MCM 1983].

2.3.2. Stimulus Response

Stimulus-response learning is a relatively simple learning strategy[11] in which stimuli and responses are classified as they occur and the nature of their relationships over time are remembered to invoke similar responses in the future. Examples include supervised and unsupervised learning, defined as:

> Unsupervised learning signifies a mode of machine learning where the system is not told the "right answer" - for example, it is not trained on pairs consisting of an input and the desired output. Instead the system is given the input patterns and is left to find interesting patterns, regularities, or clusterings among them.
>
> Supervised learning is a kind of machine learning where the learning algorithm is provided with a set of inputs for the algorithm along with the corresponding correct outputs, and learning involves the algorithm comparing its current actual output with the correct or target outputs, so that it knows what its error is, and modifies things accordingly. [AAAI 2004]

2.3.3. Process Learning

Process learning is simply defined as learning from mistakes to improve subsequent processes within the integrated cognitive system's capabilities.

2.3.4. Instruction

Instruction begins where the cognitive system can assimilate knowledge through instruction provided by humans or other cognitive systems. The cognitive system may be able to apply such knowledge to future behavior within the capabilities of the system. For example, a very capable cognitive system may learn with stories with explanations.

2.3.5. Discourse

Cognitive systems with adequate ingredients will be able to hold discourse on any topic as a method of learning from human subject matter experts and other cognitive systems. Learning through discourse may be initiated by the external human or by a self-motivated cognitive system. Discourse is a higher level of learning that requires dialog and unification at or above the understanding level. Such cognitive systems need to know how to ask questions about ingredients in the knowledge abstraction axis, e.g., semantic facts or episodic facts (e.g., procedures) that require clarification.

---

11 There are many definitions of learning mechanisms that include specific reference to knowledge representation and reasoning. For example, definitions of learning in AI are found in [Wilson 2004].

2.3.6. Exploration and Discovery

Cognitive systems with ample other ingredients may be capable of a self-motivated search for truth by exploration and/or discovery within the means of the cognitive system and its associated team members.

2.4. Reasoning Domains

The INCOG framework provides an ordered list of general purpose domains for reasoning. Included in the challenges of reasoning domains is that of representing information in the domain as well as formulating and implementing reasoning algorithms for the domain. It can be argued that these challenges increase as we move outward on the reasoning domains axis of the framework. Spatial reasoning, for example, can be considered as more complex than temporal reasoning since it commonly involves up to three dimensions, where time ordinarily has but one. Causality may be considered more complex still, as it often involves both temporal and spatial dimensions. And commonsense reasoning may be seen as involving all the prior reasoning domains (time, space, causality, and resources). The other categories of this dimension, however, are less readily compared. So, any ordering of their levels of difficulty will be more controversial.

2.4.1. Time

Reasoning about time includes all of the well-known problems of temporal reasoning. Both qualitative reasoning about orderings in time and quantitative reasoning about durations and continuously changing properties are within the scope of human and machine temporal reasoning. Such reasoning may utilize precise temporal dates, times, durations, and/or equations; or may involve temporal uncertainty, not just qualitative temporal ordering uncertainties, but alternative (disjunctive) uncertainties, as in alternate dates on which an event may occur or have occurred. Temporal reasoning may be couched in terms of time points, time intervals, or both. Reasoning with time may involve temporal constraint propagation for planning or scheduling; temporal "projection" of known facts into the future; or solving time-dependent equations of continuously variable quantities.

2.4.2. Space

Reasoning about space covers single to multi-dimensional spatial reasoning, including modeling the spatial properties of 3-dimensional objects and their relationships with each other. Subcategories of spatial reasoning include 3-D modeling of physical objects, model-based image understanding, path planning, parts-assembly planning, and automated design of spatially structured systems and components thereof.

2.4.3. Causality

Reasoning involving causality includes simple qualitative "one-shot" causation of a discrete effect by its cause (e.g., in object collisions), as well as the application of more complex causal laws and equations governing physical processes, such as the $F=ma$ of

Newtonian physics or the $E=mc^2$ of relativity. Causal reasoning plays a central role in planning and prediction, where the causal consequences of actions and events are inferred. It is also central to diagnostic reasoning, where the consequences (or symptoms) are known while triggering events and/or other causal conditions are inferred. Causal reasoning is at the heart of scientific investigation and discovery, where the aim is not the application of causal laws, but their discovery.

2.4.4. Resources

Reasoning about resources is primarily concerned with resource-constrained planning and scheduling. This includes planning with consumable resources, such as fuel, ammunition, and energy; as well as with relatively fixed resources, such as real estate and infrastructure. Of special concern is time-constrained reasoning in which time itself is treated as a resource whose consumption by planning itself constrains the whole planning process. Resource reasoning may be as simple as assessing the availability of sufficient resources in the preconditions of a single planned action, or as complicated as scheduling the utilization of multiple spacecraft instruments for gathering data in planetary flybys where resources of electrical power, maneuvering fuel, working instruments, spacecraft attitude, viewing time-windows, computer memory, and planning time must all be taken into account.

2.4.5. Commonsense

Commonsense reasoning covers a broad range of topics that fall under the relatively vague ordinary concept of commonsense. One area that has received considerable research attention is that of commonsense physics, which seeks to reason with the physical models and laws used in understanding the interactions of objects and substances in everyday life.[12] Another area is commonsense (or folk) psychology, which attempts to model common conceptions of human psychology, including psychological states, such as knowledge, belief, and anger; and psychological architectures, such as the popular Belief-Desire-Intention (BDI) model widely used in software agents.[13] Other research in commonsense reasoning focuses on characteristic commonsense reasoning patterns, such as "plausible reasoning," which has been modeled by nonmonotonic logics, probabilistic systems, and assumption-based truth maintenance.

2.4.6. Preferences

Reasoning about preferences among alternative states of affairs or outcomes of events may be used as a guide to reasoning about the most effective planning and execution for achieving preferred conditions.

---

[12] Early classic works in AI on commonsense physics include [HOBMOR 1985] and [BOBROW 1985].

[13] The foundations of the BDI model are described in [BRTMN 1987].

2.4.7. Purpose and Taboos

High-level purposes, such as personal survival or knowledge acquisition, and taboos, such as those against killing sentient beings, may also serve to guide planning to achieve more specific goals. These constraints can help ensure that ultimate purposes are promoted, and violation of taboos is avoided. Agents with integrated cognition ought to be capable of accommodating such purposes and taboos in their planning and execution.

2.5. Reasoning Mechanisms

Reasoning mechanisms are systems for generating inferences or conclusions, based on premises. This category includes both aspects of specification of information in premises and conclusions, as well as its associated semantics, inference rules, and algorithms for generating conclusions. Probabilistic reasoning, for example, requires specification with numerical qualifiers to indicate the probabilities of premises and conclusions.

Different types of reasoning mechanisms vary in the relations supported between premises and conclusions. With valid inferences in deductive logic, the premises entail the content of conclusions, while other forms of inference are weaker in that the content of a conclusion could be false even when all the premises involved are true.

2.5.1. Deductive

Deductive reasoning is a branch of cognitive psychology investigating systems ability to recognize a special relation between statements. Deductive logic is a branch of philosophy and mathematics investigating the same relation. We can call this relation entailment, and it holds between a set of statements (the premises) and a further statement (the conclusion) if the conclusion must be true whenever all the premises are true.[14]

2.5.2. Nonmonotonic

Nonmonotonic logics are used to formalize plausible reasoning. They allow more general reasoning than standard deductive logics, which deal with universal exceptionless inferences. We quote Professor Minker of the University of Maryland:

> *The subject matter of nonmonotonic reasoning is that of developing reasoning systems that model the way in which commonsense is used by humans. Nonmonotonic reasoning must therefore be able to leap to conclusions and be sufficiently robust so that when a conclusion reached by nonmonotonic reasoning is shown to be wrong it may be revised. Nonmonotonic reasoning is based on classical logic but it is a new logic developed exclusively by workers in artificial intelligence It is a significant departure from the views of logicians and philosophers concerning humans and reasoning. [MNKER 1993]*

[14] This is a cognitive system generalization of a cognitive psychology definition, see e.g., [WILKEL 2001].

Formally, the theorems of a theory in a nonmonotonic logic need not increase monotonically (i.e., with only positive changes) with the addition of new axioms. This allows exceptions to rules to be added as axioms that “defeat” or remove prior conclusions. Nonmonotonic reasoning has been applied widely to the default inheritance of properties of classes of objects by their subclasses, and to default causal inferences for expected consequences in planning actions.

2.5.3. Probabilistic

Probabilistic reasoning is the formation of probability judgments and of subjective beliefs about the likelihood of outcomes and the frequencies of events [WILKEL 2001]. Although the probability calculus can also be applied to inductive and abductive reasoning, we distinguish these as separate categories of reasoning.

2.5.4. Inductive

Inductive reasoning is reasoning from facts to a generalization about them. Inductive reasoning may infer simple empirical generalizations, for example that all objects of a certain type share a property that has been observed in all observational instances of that object type (e.g., all crows are black). Elaborate scientific theories may also be based on inductive generalizations (perhaps generalizing on empirical generalizations). Developing scientific theories, however, is widely recognized as involving other types of reasoning as well, such as application of “Occam’s razor” and conformance with other theories.

Induction is one kind of inference that introduces uncertainty, in contrast to Deductive reasoning in which the truth of a conclusion follows necessarily from the truth of the premises.

2.5.5. Abductive

Abductive reasoning is reasoning in which explanatory hypotheses are formed and evaluated. Diagnosis of the causes of the manifestations of some disorder (e.g., of equipment malfunction or disease) is a paradigm of abductive reasoning. An adequate formalization would have to take into account the following aspects of abduction: explanation is not deduction; hypotheses are layered; abduction is sometimes creative; hypotheses may be revolutionary; completeness is elusive; simplicity is complex; and Abductive reasoning may be visual and non-sentential [THGSHL 1997]. Abductive reasoning seeks the “best” explanation of a situation based on limited knowledge.

2.5.6. Analogical

Analogy is 1) similarity in which the same relations hold between different domains or systems; 2) inference that if two things agree in certain respects then they probably agree in others. A number of researchers believe this to be the strongest form of human reasoning as it supports discovery of new concepts and relationships.

2.6. Knowledge Abstraction

Knowledge is the collection of persisting memories used by cognitive systems to support reasoning, learning, and unification about the internal and external world. Internal knowledge is formulated at the self-reflective and above level of the multi-level mind. Knowledge abstraction provides the disciplined classification of knowledge in levels of increasing complexity, from simple definitional facts to complex stories and lessons.

2.6.1. Semantic Facts

Semantic facts define the meaning of terms and symbols used to express knowledge. Symbols may be linguistic or non-linguistic, including a broad range of sensory representations. Semantic facts do not cover broader relationships expressed at higher levels of knowledge abstractions, e.g., in episodic facts and stories. Facts about particular objects, events, or actors in the world are included in this category of knowledge abstraction, insofar as they independent of episodes. Ontologies, used in computer systems, express semantic facts that specify individual semantic relationships.

In humans, semantic facts are contained in semantic memory described in the MITECS Abstracts [WILKEL 2001] as:

> *Semantic memory . . . allows humans and nonhuman animals to acquire and use knowledge about their world. Although humans habitually express and exchange their knowledge through language, language is not necessary for either remembering past experiences or knowing facts about the world.*

2.6.2. Episodic Facts

Episodic facts encode previous episodic experiences as perceived by the cognitive system. The following quotations [WILKEL 2001] concerning human episodic memories should motivate the cognitive system definition of episodic facts:

> *Episodic memory is a recently evolved, late developing, past-oriented memory system, probably unique to humans, that allows remembering of previous experiences as experienced. William JAMES (1890) discussed it as simply "memory." The advent of many different forms of memory since James's time has made adjectival modifications of the term necessary.*
>
> *Our ability to remember events in our daily life and acquire specific facts after reading a newspaper or watching a newscast underscores our ability to rapidly acquire new memories. In general, these memories encode who did what to whom where and when, and have been described as episodic memories.*

2.6.3. Concepts

Concepts are abstractions that relate sets of semantic and episodic facts. Concepts are durable, possibly extensible, and discoverable. Concepts may be derived from sets of

particular facts about objects classified as similar type or may be composed from other concepts. The MITECS abstracts [WILKEL 2001] describe concepts as:

> *The elements from which propositional thought is constructed, thus providing a means of understanding the world, concepts are used to interpret our current experience by classifying it as being of a particular kind, and hence relating it to prior knowledge. The concept of "concept" is central to many of the cognitive sciences.*

2.6.4. Components

Components are related collections of episodic and semantic facts and concepts that support understanding of actors, objects, and events. An example of a component is a trace of actors, chains of events, objects or other abstractions through space and time and their interaction with other objects, events, and actors. Another example of a component is the motivation of each of the actors. Components may include: historical and situational context, alternatives, plans, consequences, and mood, etc.

2.6.5. Relations

The Relations category in our knowledge abstraction hierarchy is comprised of relations among lower-level knowledge abstractions and possible classes or instances of stories. These Relations may be used to recompose a story as a narrative along with a formalized story mark-up language.

2.6.6. Stories

The Stories category of knowledge abstraction is broader than the ordinary concept of accounts of specific incidents or events. In addition to ordinary fictional and historical stories told in books, by individuals, on stage, through movies, or other media, we include expository accounts of areas of knowledge, such as textbooks, journal articles, lectures, and the like. Such accounts share attributes of ordinary stories in that they are composed of lower-level knowledge abstractions woven together through relations to express underlying motivations, themes, and "lessons." A physics text, for example, could be understood to provide quasi-narrative accounts of how to solve physics problems, possibly accompanied by historical sketches of the development of the underlying physical theories.

2.6.7. Lessons

Lessons are what may be learned from stories to guide future behavior of cognitive systems and their offspring.

3. Cognitive Systems Survey Results

3.1. Overview

We break the existing cognitive systems into two categories: established systems and newer systems. Every cognitive system surveyed could potentially contribute concepts, designs, and built components to a new integrated cognition architectures.[15] Review of each surveyed system was conducted by the researchers who developed and extended the architectures. This was done though a series of discussions with each research team that culminated with their submission to the authors of the final charts included in this paper.

We make the following general observations about the cognitive systems surveyed:

1. Working implementations tend to require significant low-level programming for ingredients not already present in the architectural models, e.g., SOAR, ACT-R, dMARS, UMPRS, …

2. Learning capabilities are bounded to refinement within the scope of initial knowledge bases; few systems seek to understand, or invoke learning strategies beyond process learning related to current knowledge.

3. Current systems tend to be weak with respect to self-reflection, and knowledge sharing and consequently would be difficult to employ in a heterogeneous integrated cognitive architecture without extensions.

4. Current systems have core capabilities that cluster near the center of the strawman integrated multi-dimensional framework.

5. Newer or proposed systems have extended coverage of the strawman multi-dimensional framework to near the periphery in some dimensions.

6. Coverage is sparse within a specific ingredient dimension because of the large number of potentially interrelated reasoning domains not addressed in any of the architectures.

[15] The strawman integrated cognition framework proposed within this paper describes elements of an extensive concept for cognition architectures.

7. There is a significant set of advanced components (established cognitive systems) that could be included in integrated architectures that together would provide significant new capability for cognitive systems.

3.2. Established Systems

Established systems have some maturity and have demonstrated some capability in a variety of integrated cognition dimensions, and have a respectable user community.

Established systems include:

1. SOAR
2. ACT-R
3. dMARS
4. ICARUS
5. DARWIN
6. UMPRS

3.2.1. SOAR – University of Michigan – Laird

SOAR is a general purpose architecture designed as an unified theory of cognition by John Laird, Paul Rosenbloom, and Allen Newell [RLN 1993]. It is a production rule system based on the simple decision cycle - elaboration of state, choice of operators, selection of operator, and actions. Soar has a relatively large user base among existing cognitive architectures. It is supported by the University of Michigan and has been applied commercially by Soar Technology Inc. Input for assessment of Soar's support for the elements of the INCOG framework (Figure 3 of Appendix A[16]) was provided by John Laird of the University of Michigan and Robert Wray of Soar Technology.

Utility: Soar is embeddable with extension and rule sets to implement many components of an integrated cognition system.

3.2.2. ACT-R – University of Pennsylvania – Anderson, Lebiere

ACT-R [ANDLE 1998] is a cognitive architecture using production rules developed at Carnegie Mellon University (CMU) by John Anderson and Christian Lebiere. It includes a detailed approach to integrating multiple modules that correspond to different cognitive functions. The fundamental controlling structure in cognition is reactive–where production rules respond to patterns of information in various cognitive buffers.

---

16 All Figure references in the remainder of this section are contained in Appendix A.

Successive versions of ACT-R have seen widespread applications to problems of cognitive and behavioral modeling. Input for assessment of ACT-R's support for the elements of the INCOG framework (Figure 4) was provided by John Anderson and Christian Lebiere of CMU.

Utility: ACT-R based systems are embeddable for many components of an integrated cognition system.

3.2.3. dMARS – Precedence Research Australia – Georgeff

The distributed Multi-Agent Reasoning System (dMARS) is a C++ implementation of an architecture based on the BDI cognitive model [IKLW 1998]. It was developed by Michael Georgeff as a more powerful successor to the Procedural Reasoning System (PRS). dMARS has been applied to a very wide range of applications, including command and control of robotics and spacecraft, and situation awareness for the Australian Defense Forces. Input for assessment of dMAR's coverage of the INCOG framework (Figure 5) was provided by Michael Georgeff of Georgeff Inc.

Utility: BDI concepts may be useful for integrated cognitive systems.

3.2.4. ICARUS – Stanford University – Langley

ICARUS is an architecture for intelligent agents developed by Dan Shapiro and Pat Langley of the Center for the Study of Language and Information at Stanford University [LSAS 2002]. ICARUS is distinguished by its incorporation of affective values into memory and behavior; the primacy of categorization over execution and of execution over problem solving; and the internal determination of tasks, intentions, and rewards. Input for assessment of ICARUS's coverage of the INCOG framework (Figure 6) was provided by Pat Langley of Stanford University.

Agents in ICARUS incorporate long-term memory of hierarchical skills that encode how to accomplish objectives and a long-term memory of hierarchical concepts that describe how to recognize states of the environment. Both skills and concepts have associated value functions that the architecture uses to control its behavior. The system also includes short-term memories for intentions and beliefs, each element of which has associated affective values. The ICARUS focus is on reactive tasks that can be managed with the knowledge expressed with conceptual goals.

Utility: ICARUS is a general purpose architecture which is not yet broadly supported. Potential applications to integrated cognition remain to be determined.

3.2.5. DARWIN - NSI – Edelman

DARWIN refers to a series of implementations of large-scale (over 50,000 cells and 600,000 synapses) synthetic models of neural structures supporting the evolution of pattern recognition and sensorimotor coordination in a synthetic environment [RSE 1990]. It

has been developed by Reeke, Sporns, and Edelman, of the Neurosciences Institute and Rockefeller University based on Edelman's theories of Neural Darwinism [EDEL 1987]. Input for assessment of DARWIN's coverage of the INCOG framework (Figure 7) was provided by Gerald Edelman of the Neurosciences Institute.

DARWIN uses synthetic neural modeling for a multi-level theoretical approach to problem of understanding neuronal bases of adaptive behavior. Models include the environment and large-scale models of neurons and control mechanisms. Applications included automata for pattern recognition.

Utility: Darwin could provide real time sensor steering and target tracking components for integrated cognition.

3.2.6. UMPRS – University of Michigan – Huber

UMPRS (the University of Michigan implementation of PRS) is a general purpose implementation of the PRS [LHKD 1994]. It does not provide (i.e., “impose”) specific capabilities or representations on agent programmers, but rather provides a framework for their implementation. Hence, its core capabilities cover relatively few of the INCOG framework ingredients, although UMPRS applications have covered many more. Unification in UMPRS is focused on goals and planning and not reactive tasks. Input for assessment of UMPRS's coverage of the INCOG framework (Figure 8) was provided by Marcus J. Huber of Intelligent Reasoning Systems.

Hypothesis: UMPRS provides useful concepts for fully capable integrated cognition systems.

3.3. New Systems

New systems have new integration strategies and mechanisms with respect to the established systems category previously described. Further, the capabilities may not be yet implemented, and generally lack a large using community.

New systems include:

1. Shruti Smirti
2. LISA SAGE
3. Panalogy Architecture
4. Novamente
5. JAM - IRS
6. Daydreamer
7. SME, SEQL and MAC/FAC
8. ThoughtTreasure
9. Stigmeric Cognition

3.3.1. Shruti/Smirti – UC Berkeley – Shastri

Shruti and Smirti are related architectures developed by Lokendra Shastri of UC Berkeley [SHAST 1999]. They demonstrate how simple, neuron-like, elements can encode a large body of relational causal knowledge and provide a basis for reactive, rapid inference. Input for assessment of Shruti/Smirti's coverage of the INCOG framework (Figure 9) was provided by Lokendra Shastri of UC Berkeley.

Utility: Shruti provides a key cognitive real-time component that supports reactive text understanding. It may provide a general model for composition for integrated cognition systems.

3.3.2. SAGE – HRL, UCLA, Furmanski, Hummel, Holyoak

SAGE (Self-Aware Adaptive Generalization Engine) is a cognitive-based architecture that is adaptive, self-reliant, and can reason by analogy (like people do) in order to discover meaningful relationships between seemingly dissimilar data. It blends connectionist/neural networks with symbolic systems. Its self-supervised learning uses self-reflective algorithms that allow the system to acquire new knowledge, learn from its past, and avoid extensive human intervention by guiding its own performance. Applications include roles a as network security watch dog, decision aid for intelligence, or strategic agent for military simulation. Input for assessment of SAGE's coverage of the INCOG framework (**Error! Reference source not found.**) was provided by Chris Furmanski and John Hummel of HRL Laboratories, LLC, and UCLA.

Utility: A very general integrated model that provides composition concepts and components for integrated cognition systems.

3.3.3. Panalogy Architecture - MIT - Singh, Minsky

Panalogy is a cognitive architecture designed to support commonsense reasoning being developed by Push Singh and Marvin Minsky of MIT [SINGH, 2003]. It utilizes multiple strategies for representation and reasoning to support typical types of commonsense reasoning. It has a capability to learn how to learn, includes commonsense understanding, and more. Input for assessment of Panalogy's coverage of the INCOG framework (Figure 11) was provided by Push Singh of MIT.

Utility: A very general integrated model that provides composition concepts and components for integrated cognition systems.

3.3.4. Novamente - Novamente LLC – Goerztel

Novamente is a system organized with distributed atoms of knowledge that may be employed in an unlimited number of contexts [GPSML 2003]. Atoms have truth value and attention value. Mind agents operate on these atoms, learning how to learn. Input for as-

sessment of Novamente's coverage of the INCOG framework (Figure 12) was provided by Ben Goertzel of Artificial General Intelligence Research Institute (AGIRI).

Utility: Novamente provides a very general integrated model that provides composition concepts and components for integrated cognition systems.

3.3.5. JAM – IRS – Huber

JAM is another version of the PRS [HUBER 1999]. It provides *Reactive* and *Deliberative* models as well as *Reflective* and *Self-Reflective* capabilities in the form goal semantics and meta-level reasoning. Its unification is limited to planning and goal driven behavior, and does not support much in the way of learning. Input for assessment of JAM's coverage of the INCOG framework (Figure 13) was provided by Marcus Huber of Intelligent Reasoning Systems.

Utility: The JAM model provides composition concepts and components for integrated cognition systems.

3.3.6. Daydreamer – IBM – Mueller

Daydreamer was developed by Erik Mueller of IBM to simulate a human stream of thought and its triggering and direction by emotions [MUEL 1998]. Input for assessment of Daydreamer's coverage of the INCOG framework (Figure 14) was provided by Erik Mueller of IBM.

Utility: Daydreamer provides cognitive components for reactive, affective mood management including social interaction with humans and other machines with affective ingredients of integrated cognition systems.

3.3.7. SME, SEQL, and MAC/FAC – Forbus/Gentner

A number of tools for structure-mapping and qualitative reasoning have been developed at Northwestern University by Ken Forbus, Dedre Gentner and others [FFG 1994]. These include the tools the Structure-Mapping Engine (SME), SEQL and MAC/FAC, which are assessed together in Figure 15. Input for assessment of coverage of the INCOG framework by these tools was provided by Ken Forbus of Northwestern University.

Utility: These tools provide concept and design for cognitive components for analogical reasoning for integrated cognition systems.

3.3.8. ThoughtTreasure – IBM – Mueller

ThoughtTreasure is a story understanding and commonsense reasoning system developed by Erik Mueller of IBM [MUEL 1998]. Input for assessment of ThoughtTreasure's coverage of the INCOG framework (Figure 16) was provided by Erik Mueller.

Utility: ThoughtTreasure provides concept, design, and components for commonsense evaluation of discourse for integrated cognition systems.

3.3.9. Stigmergic Cognition – Altarum – Van Parunak, Brueckner

Stigmergic Cognition consists of cognitive components that exhibit emergent behavior and have performed well in many roles [PARBRU 2003]. Input for assessment of Stigmergic Cognition's coverage of the INCOG framework (Figure 17) was provided by H. Van Parunak and Sven Brueckner of Altarum.

Utility: Stigmergic Cognition provides concept and design, and cognitive components for social behavior with goal orientation for integrated cognition systems.

4. Integrated Cognition Architectures

4.1. Use of Cognitive Components in Integrated Cognition

Some of the cognitive systems described above could be extended for integration into an integrated cognition framework and architecture and used as components themselves in such a framework. We have illustrated that notion in Figure 2.

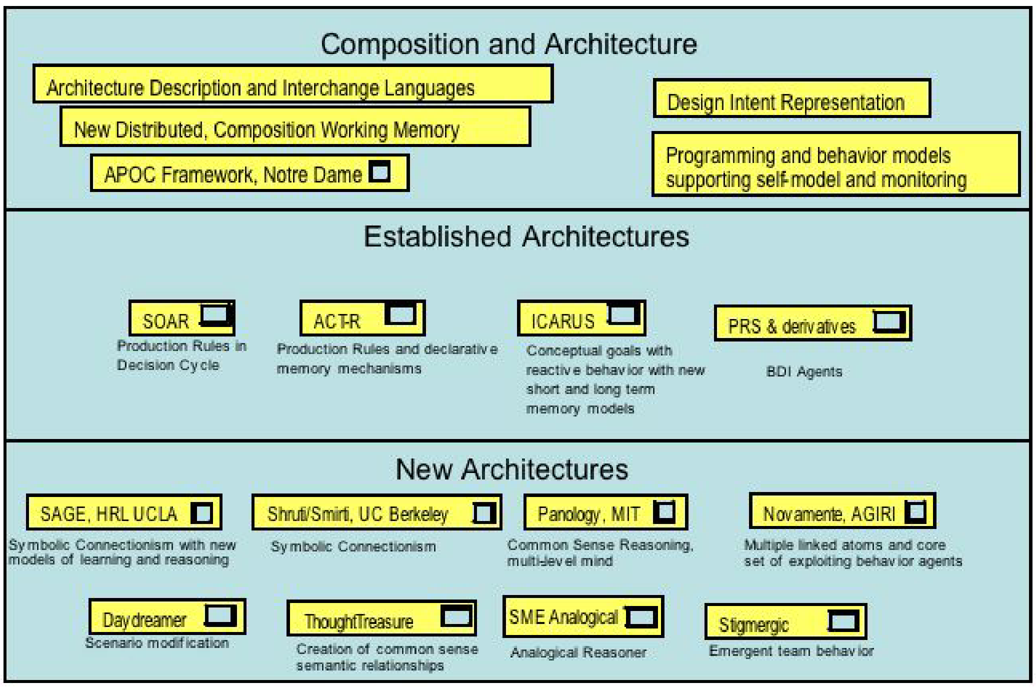

Figure 2. Example Technologies for Integrated Cognition

The complete summary of each component's capabilities as currently specified is provided in Appendix A. This appendix contains summary charts for each established and new architecture indicating the scope of capability. These architectures all lack the necessary architectural mechanisms to operate within the total scope of the strawman integrated cognition framework. This is because the components themselves do not have the spectrum of interfaces and functions necessary for an integrated cognition system as described in the strawman framework.

4.2. Example Top Level Architecture for the Strawman Framework

DARPA Information Processing Technology Office (IPTO) Deputy Director, Dr. Barbara Yoon, asked Dr. Robert Rolfe to build an illustration of a top level architecture that could serve as an example of how one might approach constructing an integrated cognition system. Further, such a top-level architecture provides the mechanisms to parcel out program work packages to achieve integrated cognition.

In Appendix B we describe such an example architecture for integrated cognition compliant with the strawman framework presented in this paper. This example of a top-level architecture for integrated cognition is defined in terms of some 64 functions, a variety of interface mechanisms, and basic architectural and design principles.

4.3. Summary

This paper set out with a goal to define a new level of integrated cognition capability in computational systems. On the way we learned that current systems exhibit a full spectrum of capabilities. As a result, we endeavored to define a new strawman framework for integrated cognition exploring the full set of relevant dimensions. We then surveyed established, those documented in the literature, and new architectures under development, and asked the respective research teams to analyze their demonstrated and projected capabilities with respect to the strawman framework. Finally, we developed an example top-level architecture to illustrate the functions, interface mechanisms, and principles necessary for a new generation of cognitive systems.

Appendix A. Survey Detailed Results

This appendix presents the summary results of the capabilities of cognitive architectures relative to the cognitive dimensions described above. Capabilities of established and new architectures are distinguished at different levels via the following definitions of color codes.

- **Green dots indicate core integrated capability**
  - Ingredient may be adequate today, but probably will require re-tooling for integrated cognition
- **Yellow dots indicate one or more example systems have been integrated with the core framework as applications**
  - This implies that there is an understanding in the community of researchers of how to interface and integrate, at least for the specific application, but perhaps not generally
  - A yellow dot alone indicates the possibility exists to add the ingredient into the existing architecture without modifying its core
- **Blue dots indicate potential extensions of the core framework**
  - There are feasible concepts to extend the core framework
  - Potentiality is reinforced by there being yellow dots as well indicating somebody has at least attempted a narrow implementation, not necessarily a general purpose or architecturally robust solution
  - Light blue dots have the potential for INCOG investment

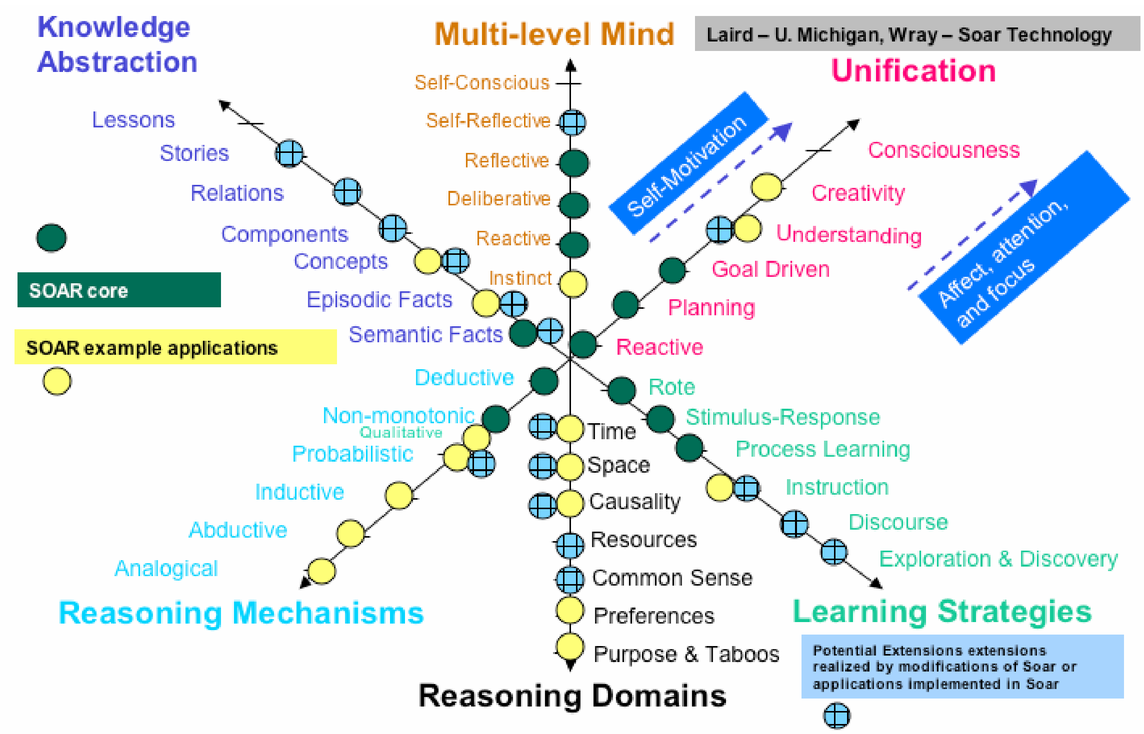

Figure 3. Soar Ingredients for Integrated Cognition

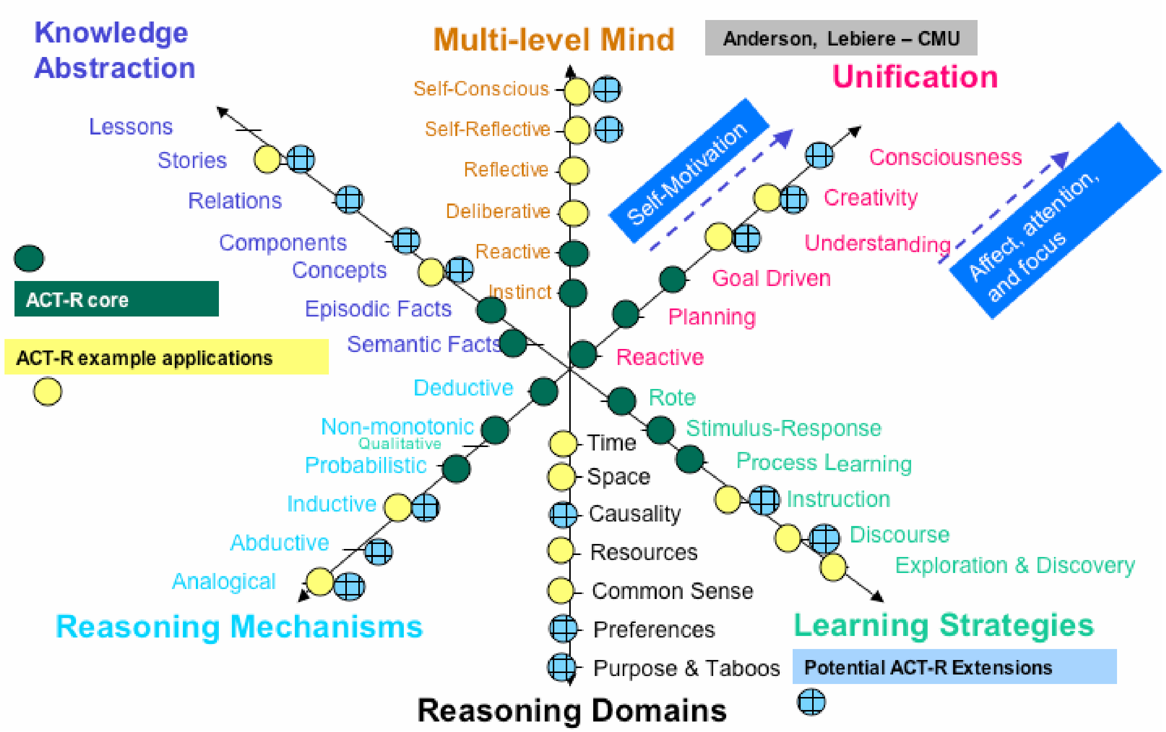

Figure 4. ACT-R Ingredients for Integrated Cognition

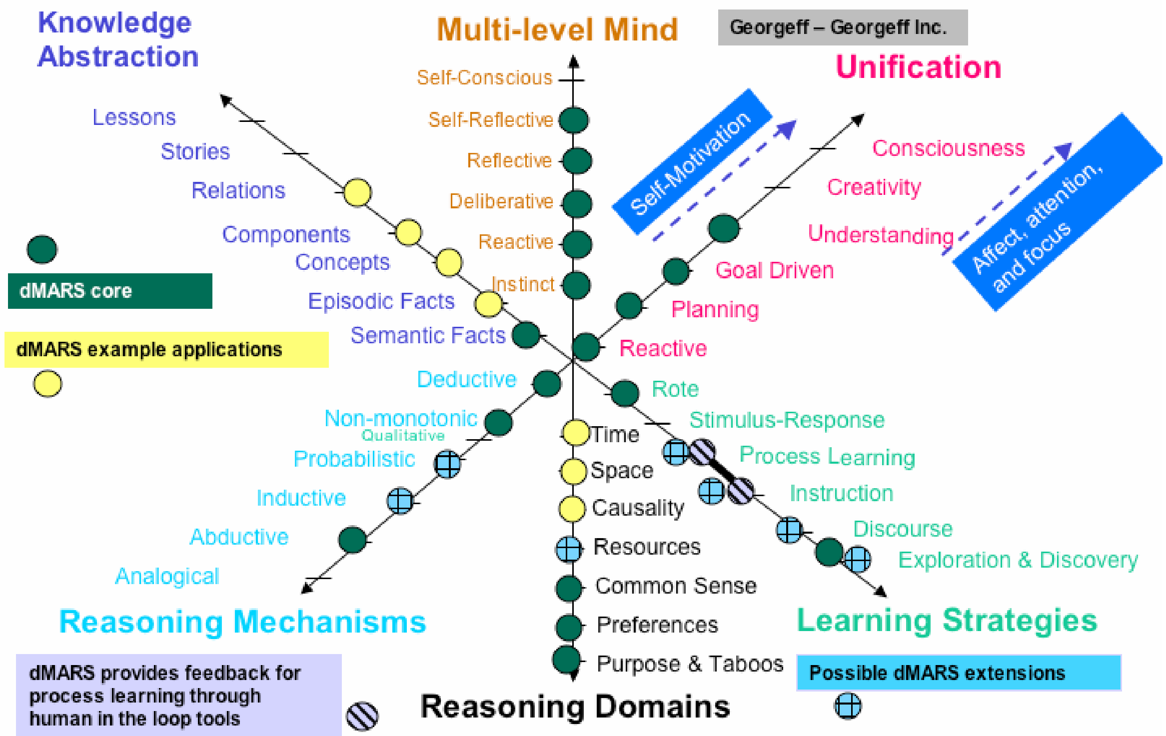

Figure 5. dMARS Ingredients for Integrated Cognition

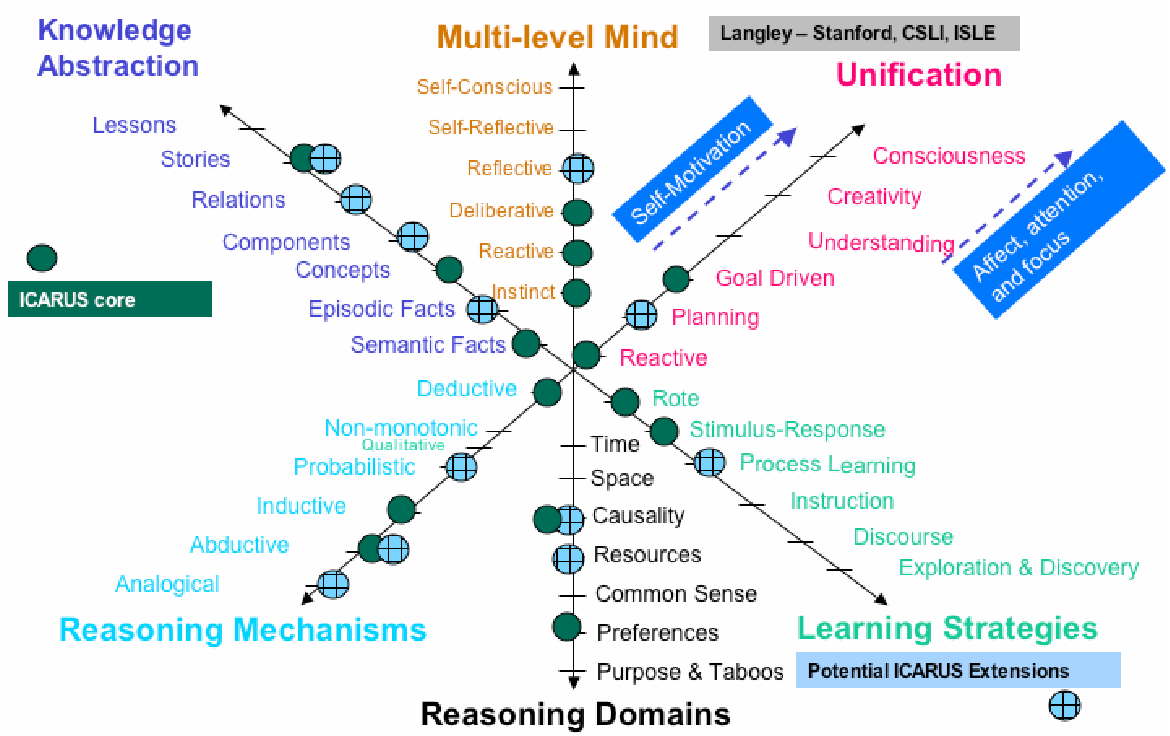

Figure 6. ICARUS Ingredients for Integrated Cognition

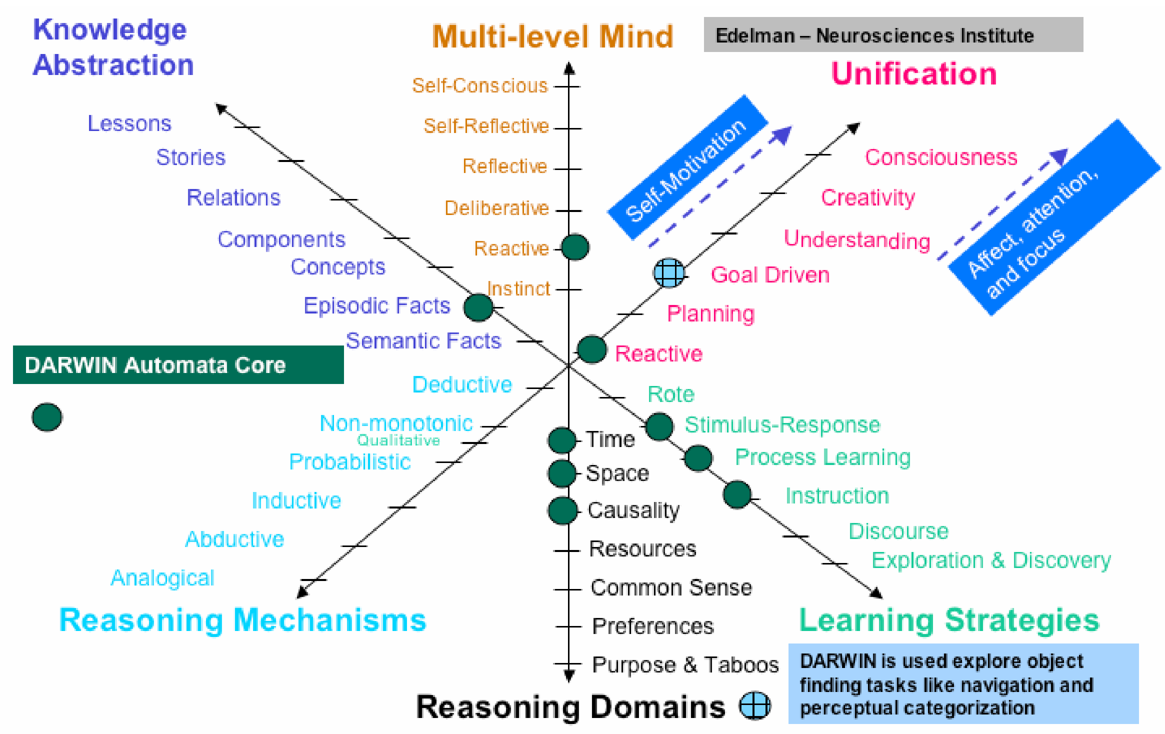

Figure 7. DARWIN Ingredients for Integrated Cognition

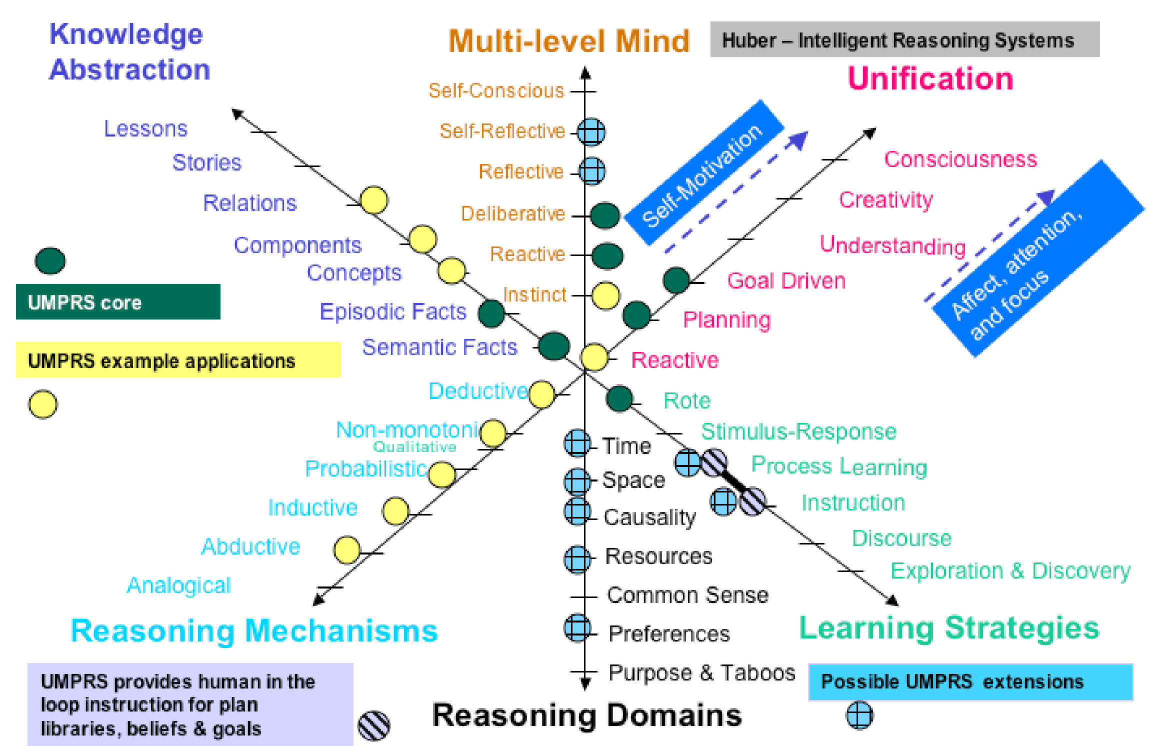

Figure 8. UMPRS Ingredients for Integrated Cognition

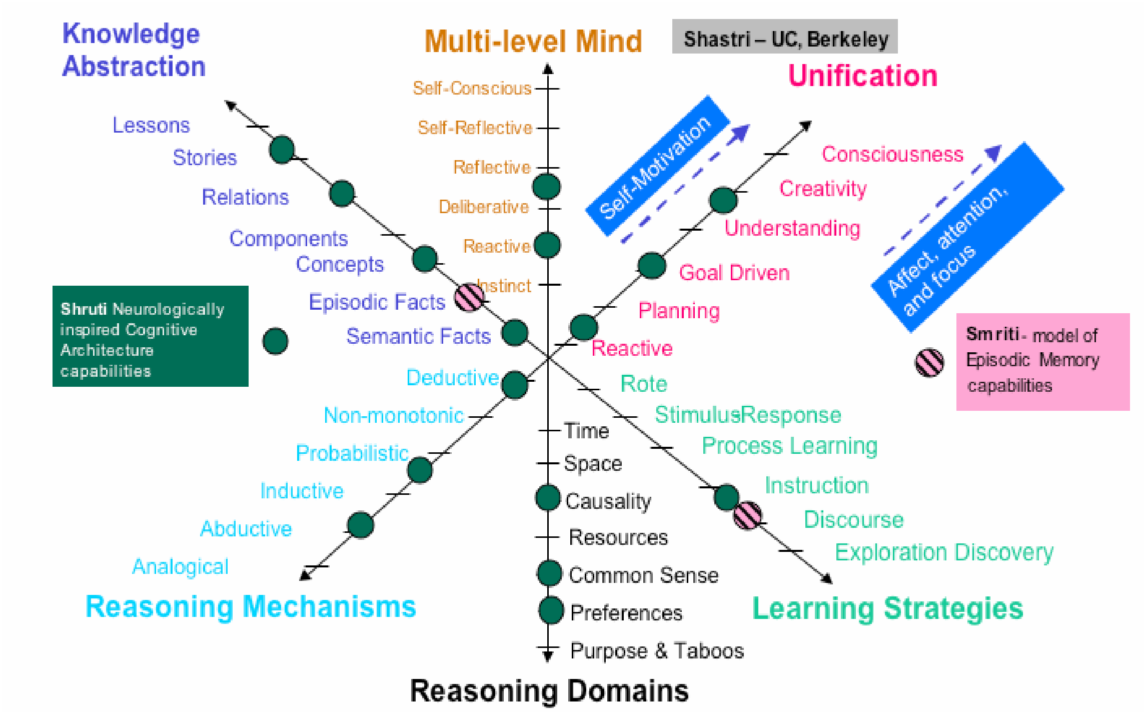

Figure 9. Shruti and Smirti Ingredients of Neurologically Inspired Cognitive Systems

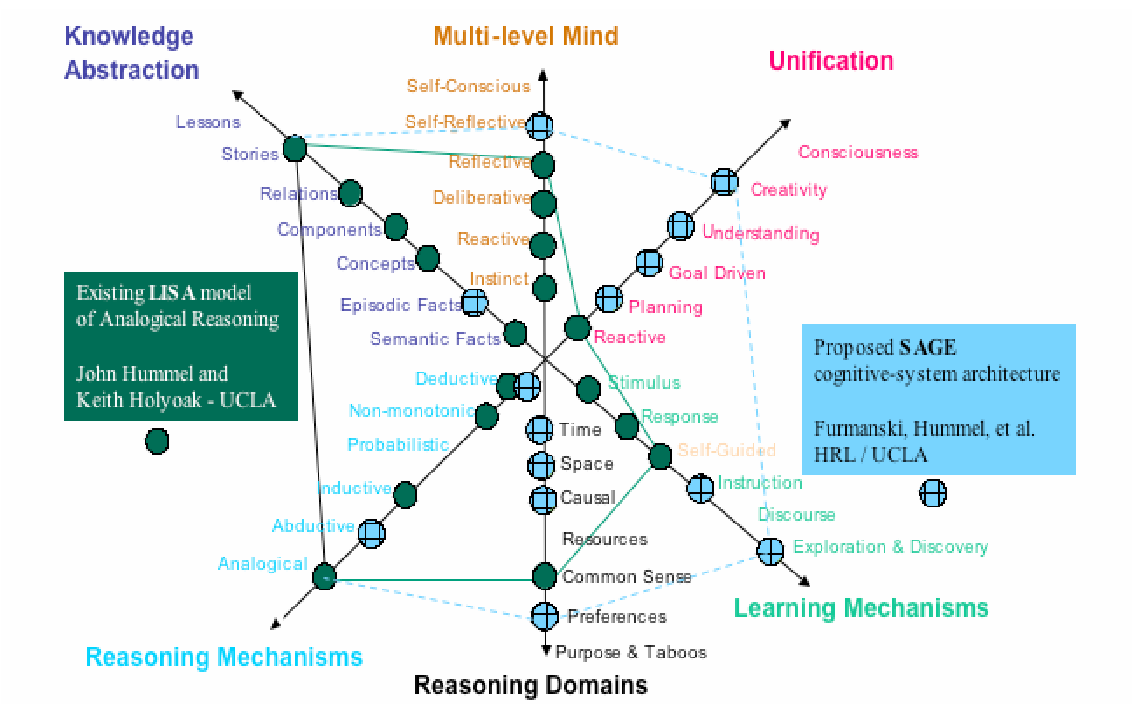

Figure 10. LISA and SAGE Ingredients of Neurologically Inspired Cognitive Systems

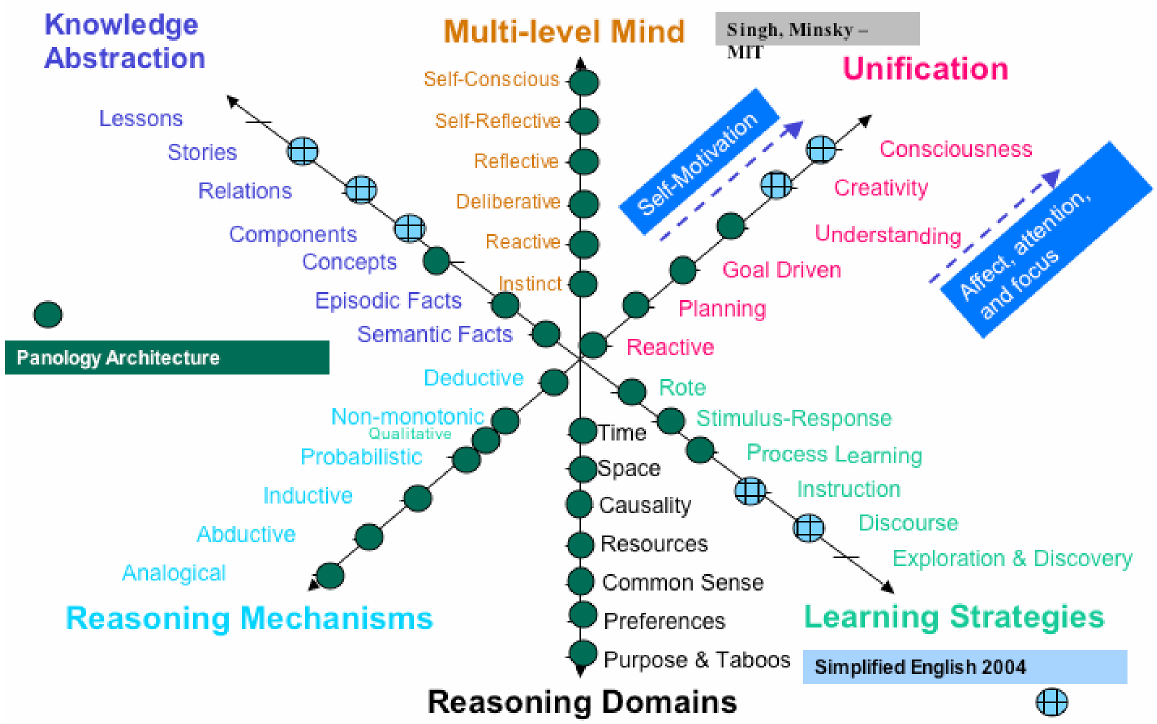

Figure 11. Panalogy Architecture Ingredients for Integrated Cognition

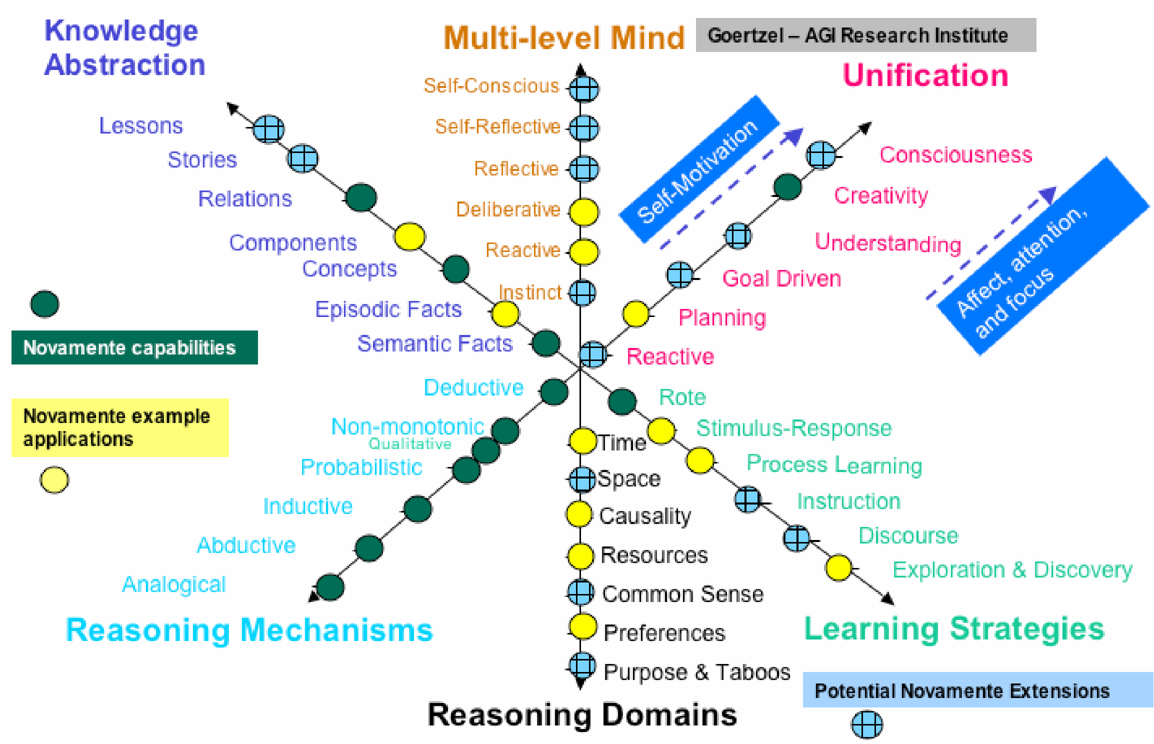

Figure 12. Novamente Ingredients for Integrated Cognition

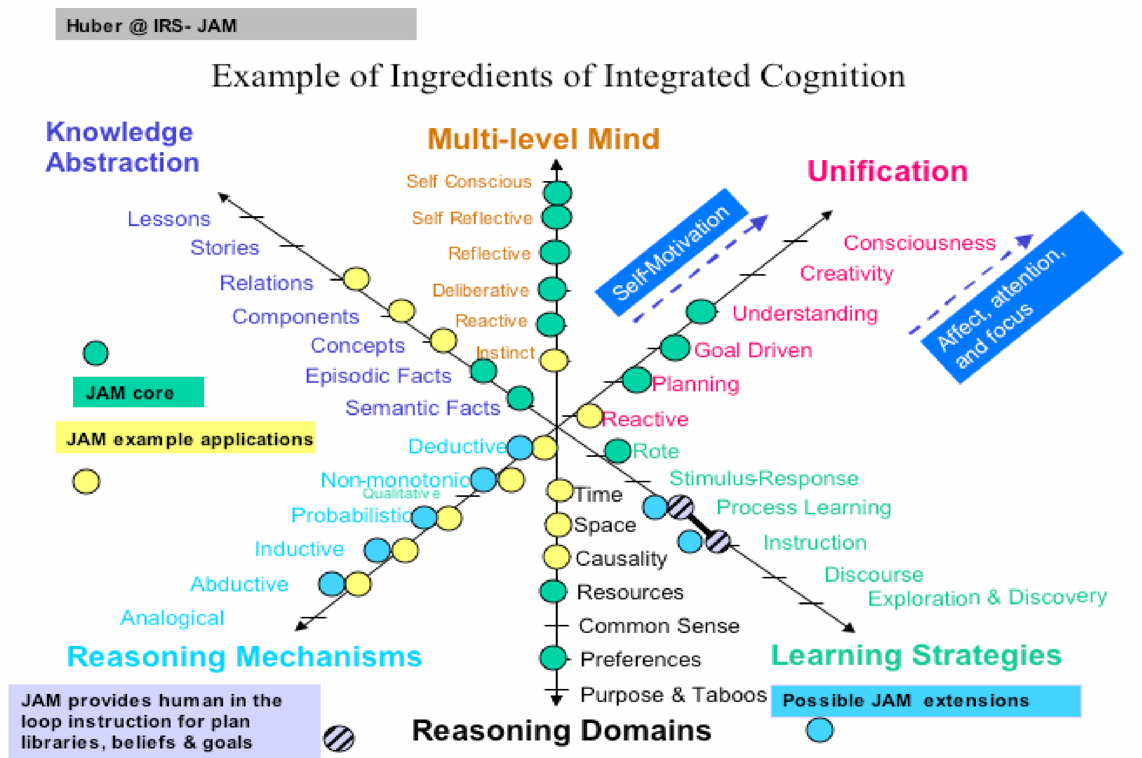

Figure 13. JAM Ingredients for Integrated Cognition

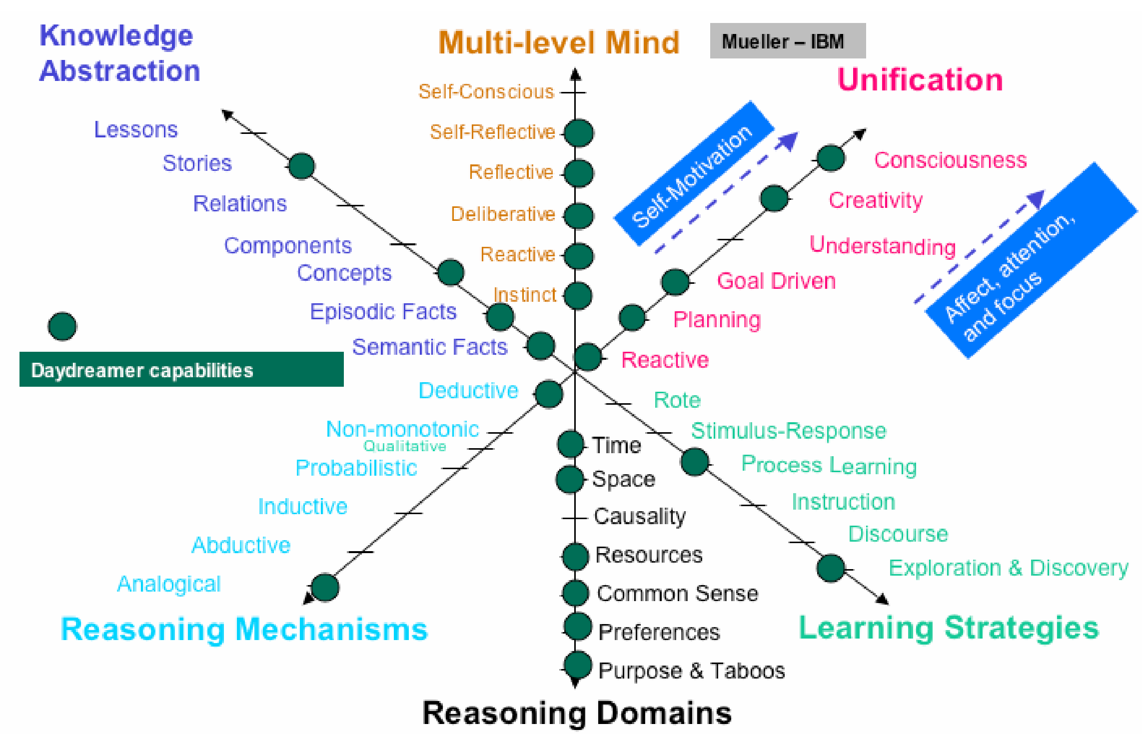

Figure 14. Daydreamer Ingredients for Integrated Cognition

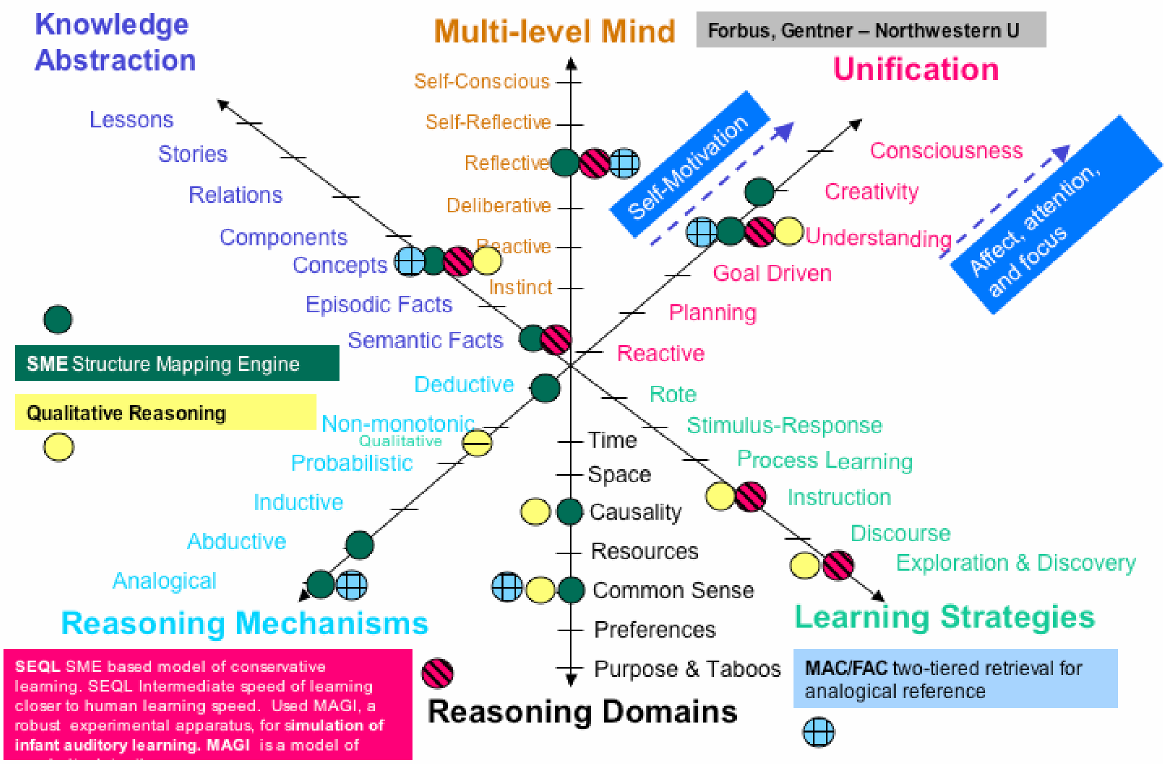

Figure 15. Structure-Mapping and Qualitative Reasoning Integrated Cognition Ingredients

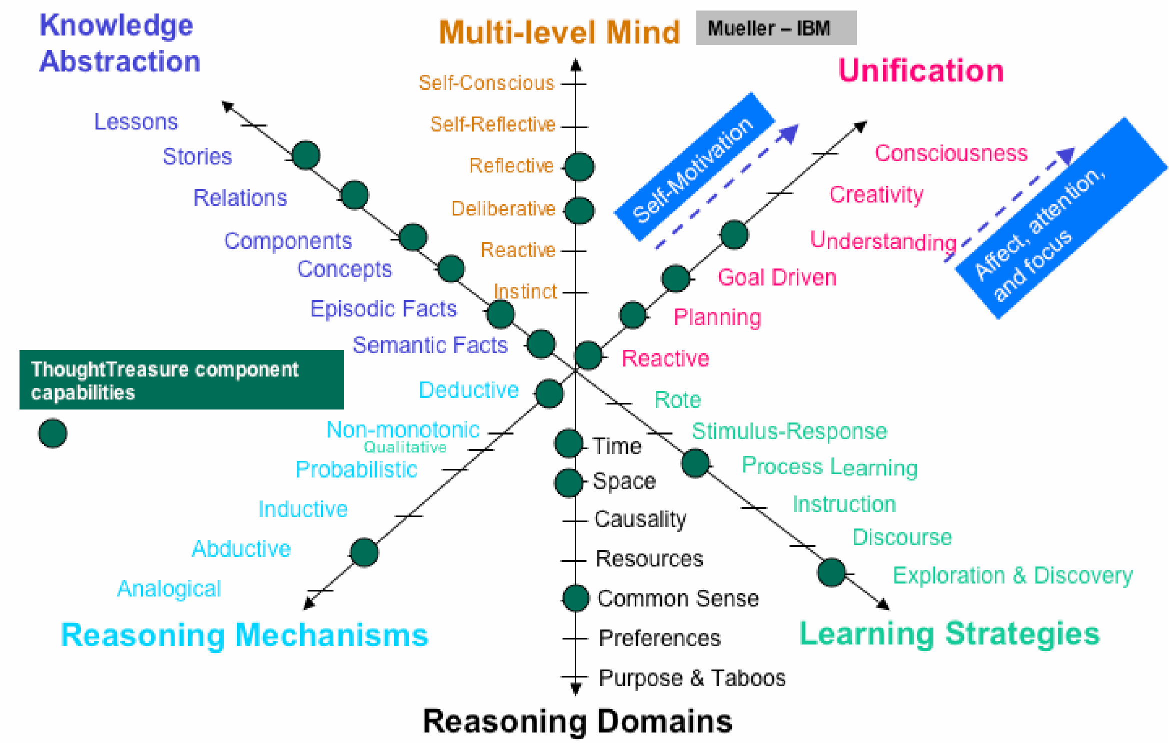

Figure 16. ThoughtTreasure Ingredients for Integrated Cognition

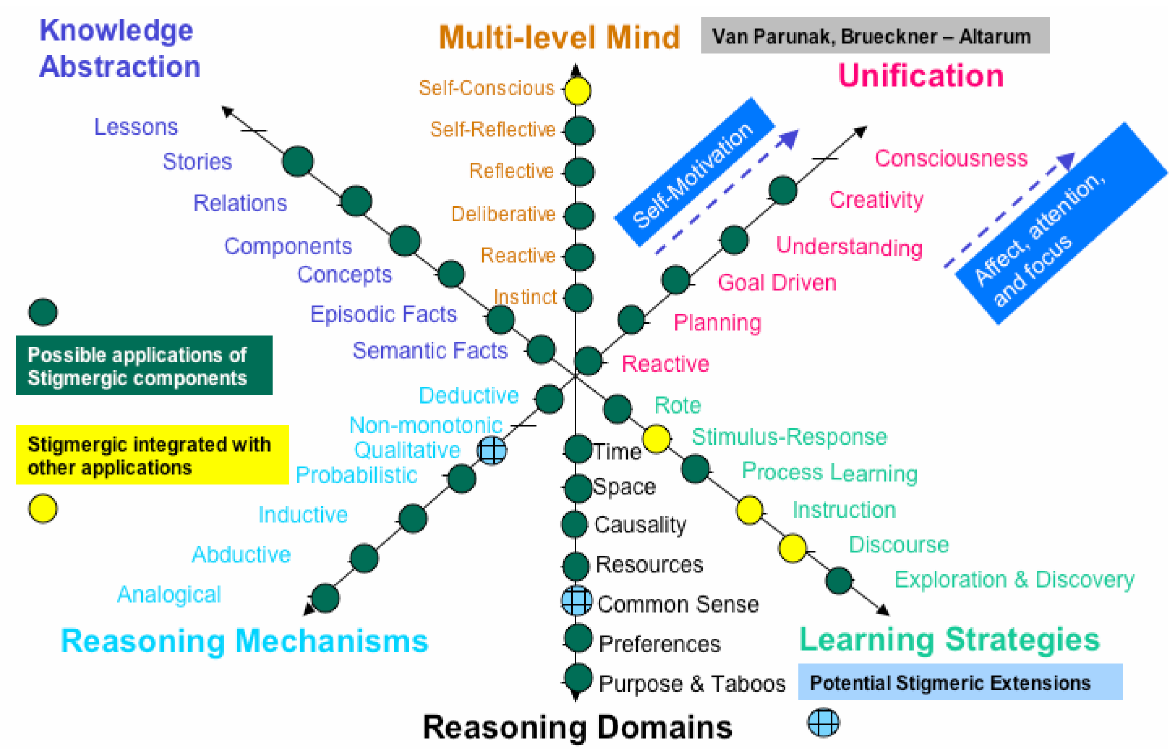

Figure 17. Stigmergic Cognition Ingredients for Integrated Cognition

Appendix B. Example Top-Level Integrated Cognition Architecture

In this appendix we provide an example top-level architecture that meets the requirements of the strawman integrated cognition framework described in this paper.

INCOG Architecture Challenge

Cognitive architectures for computers face a truly grand challenge in providing a superstructure for human-level cognitive performance. To fully meet this challenge will require providing support for many cognitive capacities, including:

- Intelligent interaction with humans and other systems that reduces the workload of the human user by anticipating and performing a variety of intelligent tasks
- A cognitive model that fuses self-consciousness, and self-reflective, deliberative, and reactive capability toward common goals and objectives on differing time horizons
- Synthesizing discourse with diverse participants
- Integrating multiple forms of reasoning and new knowledge forms
- Integrating a critical mass of knowledge to provide intelligent behavior
- Adaptation to changing objectives using a priori and learned situational knowledge
- Massive distribution and parallelism needed for real problem solving
- Self-management to dynamically reach goals and objectives

The scope of a cognitive architecture that supports everything in the cognitive framework presented above is broad, indeed. It needs to:

- Incorporate by concept and design the full scope of integrated cognition capabilities
- Support the understanding of the changing context of discourse with humans and other cognitive components

- Incorporate capabilities to operate as a team player or leader in coordination with teams of specialists with diverse sets of expertise domains

- Manage behavior with fuzzy goals, sub-goals with integrated motivation, affective and cognitive resources, a variety of models, and diverse classes of knowledge

- Be teachable by humans and cognitive systems of its class

- Be able to communicate the justification for the planned or exhibited behavior

- Manage dynamic collection of cognitive components with a high degree of parallelism necessary to reach useful results in periods dictated by external need and internal goals and capabilities

- Continuously refine and/or alter its planned hypothetical behavior until planning or urgency require new behavioral output

- Retain learned information as knowledge in symbolic and/or iconic form for future use

- Develop understanding and new learning strategies to support possible future behavior on various time horizons based on builder's imprinted values

In developing an architecture to meet these needs, many fundamental design issues need to be addressed, including an architecture for dynamic distributed processing, the design of working and long-term memory, and a variety of foundational components. Multiple interacting processes must be supported that can span as many distributed processors as possible using available technology. Working memory requires a system design that enables sharing working memory across distributed processes while supporting real-time performance needs. Long-term memory needs persistent storage of knowledge available to all processes including a variety of reasoning mechanisms. Various common foundation components are needed, which can be built once and deployed throughout the machine mind as needed, both statically and dynamically. And, substrate services need to be selected or designed for distributed computing, bootstrapping, evolution, management, persistence, cloning, and fault tolerant services.

Learning: An Example of Integration Requirements

As an example of how human-level cognition really requires integration of many cognitive ingredients or capabilities, consider learning. Full support for human-level learning capabilities can be expected to require integrating many of the ingredients of the INCOG framework, which may be grouped into the following functional areas:

- Discern / modify / validate relationships and context

    - Deductive, Nonmonotonic, Probabilistic/Inductive, Analogical, and Commonsense Reasoning Mechanisms

- Store and retrieve information at multiple levels
    - Semantic Facts, Episodic Facts, Relations, Concepts, Stories, Lessons

- Reason along multiple dimensions, e.g.,
    - Time, space, causal models, resources
    - Preferences, urgency

- Integrate learned material into behavior at the appropriate level, e.g.,
    - Reactive / instinctual
    - Planning / goal driven / Deliberative
    - Creative (novel combinations or perturbations of learned material)
    - Understanding and apparent consciousness by producing a unified picture of learned material, relationships among the material and to concept of "self," and providing explainable plans and behavior

- Acquire new knowledge, processes, tasks, and skills for multiple domains
    - Incorporate learning, understanding, assessment, and self-programming model

This is just one example among many high-level cognitive processes that require many of the ingredients of the INCOG framework in order to approach human-level cognition. Other examples will be provided below, after presenting a strawman architecture for integrated cognition.

INCOG Strawman Architecture

A top-level view of the INCOG strawman architecture is shown in Figure 18. This figure is based on Dr. Ronald Brachman's proposed cognitive architecture [BRACH 2002], which has been adapted here to better capture distinctions made in the INCOG framework. In particular, this INCOG architecture distinguishes more finely between different levels of the multi-level mind, separating Brachman's *Reactive Processes* into *Programmed Instinct* and *Learned Reactions*; and adding several other levels above the *Reflective*. This architecture also highlights discourse as a key area of human-level cognition by separately identifying it and its relations to other input and output processing.

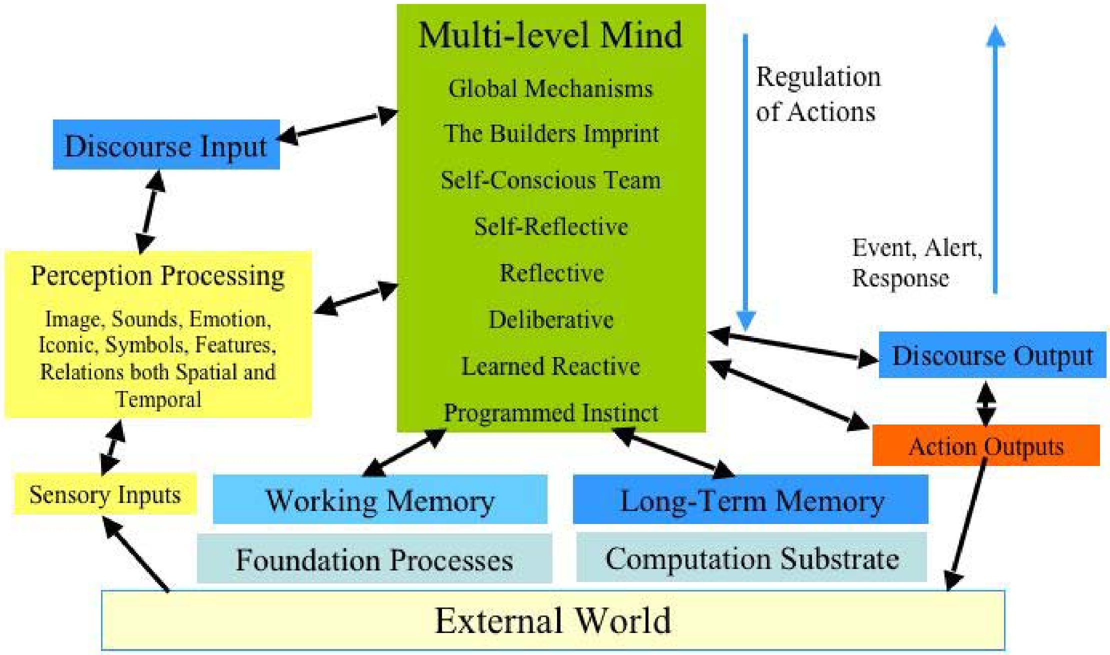

Figure 18. INCOG Strawman Architecture Top-Level Static Components

This architecture diagram represents the world external to the cognitive agent by the *External World* box at the bottom. The rest of the boxes represent information and processes of the cognitive agent. Raw *Sensory Inputs* come into the agent from its sensors and are processed initially via *Perception Processing*, which hands off linguistic data to *Discourse Input*. The results of perception and discourse input processing are fed to various levels of the *Mult-ilevel Mind*, as appropriate. The *Mult-ilevel Mind* uses inputs from *Working Memory* and *Long-Term Memory* to place new inputs in context and to help determine its responses and other activities. The *Multi-level Mind* also stores information in *Working Memory* and *Long-Term Memory* as warranted. The processing and exchange of information throughout is enabled by a host of *Foundation Processes*, as well as the *Computational Substrate*. The results of the execution of cognitive processes may then find expression in discourse via *Discourse Output* or through other *Action Outputs*. The *Discourse Outputs* consist of the intended discourse, which is communicated to the appropriate physical effectors as *Action Outputs* in order to effect speech, writing, and any other forms of communication.

In the sections that follow, different aspects of this architecture will be elaborated. First, various knowledge abstractions used in working and long-term memory will be presented. This is followed by a presentation of an extensive set of 61 functional "packages" that represent processes used by the main elements of the architecture diagrammed in Figure 18. Next, an example package description is elaborated in terms of its functional components and their relationships. Then, ontologies, which enable communication among the packages, are discussed. After that, several aspects of learning in cognitive systems are discussed, including: learning mechanisms, principles for learning, new

model capabilities for learning, linguistic learning, and learning expectations. Finally, some examples are presented of composite cognitive functions that are composed out of various combinations of the 61 functional packages. Structure charts are provided to describe the following composite cognitive functions: apparent conscious behavior, self-reflection, and speech and text processing.

Memory Knowledge Abstractions

A cognitive agent's memory must be capable of storing a wide variety of different types of information. There are many different ways of parsing this information into categories based on different criteria, such as its: reasoning domain; application domain; complexity; importance/relevance; and proximity to real-world interface via sensors or actions. Many different such parsings are relevant to the interests of organizing information appropriately for the many possible components and interfaces of cognitive systems. Here, we present just one high-level view or parsing of such information in Figure 19. This abstraction is presented at the "knowledge level"[17] to identify some of the key types of knowledge involved in the cognitive processes used within this INCOG architecture. It organizes the different types of knowledge required for the cognitive ingredients of the INCOG framework in Figure 1, with a focus on the dimensions of *Knowledge Abstractions* and *Reasoning Domains*.

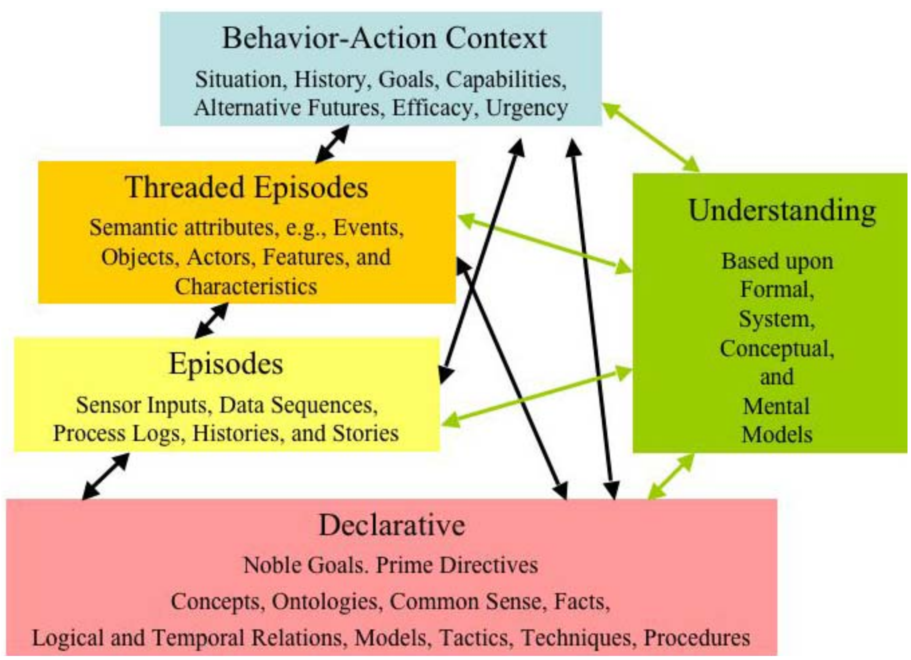

Figure 19. Long-Term Memory Knowledge Abstractions

[17] The "knowledge level" was introduced by Herbert Simon in his presidential address to the AAAI at their first conference in 1979, as recorded in [SIMON 1981, 1982] and reflected upon in [SIMON 1993].

The knowledge abstractions presented here are divided into four main groupings linked to each other and to a symbolic representation of the "understanding" of this knowledge that arises (or emerges) from it being placed within the context of different models and their associated cognitive processes. Declarative knowledge covers a broad range of information, including the "Noble Goals" or "Prime Directives of the *Builders Imprint* level of the multi-level mind. The ontologies included here capture the semantics of the vocabulary used in representing a cognitive agent's knowledge. Commonsense information may be captured in part by the ontologies and in part by other facts about various domains common to a community of cognitive agents. Other facts and their relationships–temporal, causal, and otherwise–are included, as are various models of objects and processes. Tactics, techniques, and procedures are also included as *Declarative*, insofar as they are expressed declaratively, making them amenable to manipulation by the cognitive agent, for example in selecting, assessing, and modifying them for suitable purposes.

Episodes are an abstraction of knowledge that is organized into temporal and/or causal sequences. It can be understood as being composed from *Declarative* knowledge that is organized to identify sequencing relationships. Raw sensor inputs in sequential data streams can constitute such knowledge when placed within a computational context that supports inference of the temporal relations, which may be only implicit in the data stream itself. Episodes may also be composed by explicit dating of related events or states, as in histories and some stories. Tracings of objects or processes through sequences of such episodic information constitutes knowledge of *Threaded Episodes*. A threaded episode might, for example, capture key characteristics of an actor of interest as they change over time.

The knowledge abstraction category of *Behavior-Action Context* is intended to capture knowledge about the relations between behavior or actions and their context, such as the situation and history in which they occur; the capabilities that enable them; the alternative futures that are conceived of as their possible consequences; their expected or actual efficacy in achieving related goals; and the urgency with which they are planned or performed. Such contextual knowledge may cover actual behavior and actions that have been performed, as well as possible behavior and actions that are contemplated for future action or postulated as explanations of past events. Knowledge at the level of *Behavior-Action Context* can elaborate knowledge of a *Threaded Episode* to provide, for example, the conditions that an actor found promising for an action (or series of actions) in respect of the goals of the actor and its assessment of its own capabilities.

Functional Packages for Integrated Cognition

In order to identify more of the specific functional capabilities required to instantiate this architecture, a number of functional components are described and grouped into categories corresponding to the different parts of the INCOG architecture and dimensions of the INCOG framework. This grouping of functional components into packages can be understood as analogous to the use of packages in object-oriented programming languages, such as C++ and Java. They are groupings of related functions and information. But, they

are not necessarily independent, in that functions of one package may well require use of some of the functionality found in other packages.

The package groupings presented here are the following:

- Multi-Level Mind Packages
  - Foundation Multi-Level Mechanisms
- Foundation Unification Packages
- Foundation Global Mechanisms Packages
- Foundation Reasoners Packages
- Foundation Domain Packages
- Foundation Discourse Support Packages

Two sets of packages are distinguished for the Multi-Level Mind. The first applies only to the Multi-Level Mind, while the second grouping contains foundation packages that may also be used by other parts of the INCOG architecture. The rest of these categories contain foundational packages that are represented by the *Foundational Processes* box on the top level architecture diagram of Figure 18. Such packages may be used by multiple architectural components. In the following subsections, each functional package is briefly described. A full specification of these packages is beyond the scope of this initial presentation of the architecture. Much more work will be required to develop a full design for any of these packages.

Multi-Level Mind Functional Packages

We begin the identification of individual functional packages with a set that is uniquely assigned to the Multi-Level Mind part of the cognitive architecture. These packages operate together to provide apparent conscious output that appears to mimic the single conscious state of a human. In fact, just as for the human mind, multiple independent activities will occur. Going beyond human mental capabilities, alternative futures may be explored in greater breadth and depth and rigorously evaluated by an artificial cognitive agent. And, daydreaming may be an on-going activity. Unlike embodied humans, some of the perceptual processing units may be duplicated at will to evaluate many alternatives in parallel with no loss in reactive performance.

The Multi-Level Mind packages are listed in a table (Table 1) along with brief descriptions, establishing the presentation format for all the integrated cognition packages identified here.

Table 1. Multi-Level Mind Packages

Package Name	Package Description
1. Behavior Adjudication and Unification	Nominates cognitive components and adjudicates redaction of processes, working memory, and knowledge. Resolves priorities that cannot be settled in other processes, with rules and precedence.
2. Motivation	Provides and maintain motivation, generic goals, and objectives for current state, both external and internal.
3. Affect	Provides and maintains affective state and influence affective state of others.
4. The Builders Imprint	Applies imprinted values and taboos through critics and censors to machine's functions generating visible behavior.
5. Self-Conscious	Generates adaptive behavior associated with cultural, social, and team behavior.
6. Self-Reflective	Applies self-reflective processes utilizing representations of self-descriptions, deliberations, internal states, and the external situation to evaluate capabilities and past behavior to develop future alternatives.
7. Reflective	Applies reflective processes related to executing current goals and objectives. Operates with representations of deliberations, internal states, and the external situation to evaluate available future behavior.
8. Deliberative	Applies deliberative processes related to current goals and objectives. Deliberative operates on representations of internal states, external situations, and current capabilities.
9. Learned Reactive	Applies learned and evolved reactive processes. Operate on representations of world situations with little influence from internal state except for resource conservation.
10. Programmed Instinct	Applies built-in core behavior and skills that only change though software updates.

Foundational processes used by the Multi-Level Mind and other parts of the INCOG architecture are presented in Table 2.

Table 2. Foundation Multi-Level Mind Packages

Package Name	Package Description
11. Self-Model Builder and Maintainer	Creates and maintains a dynamic model of the current working memory (or alternative) and attributes, processes, status and state models, ownership, sharing, status, and projected changes of key attributes.
12. Self-Conscious Emotion Generator	Implements, e.g., an Ortony-like model of preferences [ORTNY 2003] that are related to events, actors, and objects that are always context dependent.
13. Self-Reflective Observer	Monitors internal behavior of machine and its components and relates to temporal progress toward long-term and current goals, objectives, plans, tasks, and skills execution.
14. Reflection Tracer	Constructs an episodic thread of internal and external events for use by any process to evaluate performance.
15. Cognitive Component Performance-Evaluator	Evaluates the present and past performance of cognitive components in context of goals and objectives, internal and external states.
16. State Transition Preference Generator	Generates new preferences to create a shift in machine behavior to achieve a new outcome consistent with present goals and objectives.
17. Goals and Objectives	Manages goals and objectives stack for long-term and short-term guidance of machine behavior.
18. Prediction and Virtual Planning	Manages processes that evaluate alternative futures given real or synthetic situation assessment, capabilities, goals, objectives.
19. Alternative Future Generator	Evaluates alternative futures given a real or synthetic situation, goals, objectives, and current, increased or augmented capabilities.
20. Recommended Alternatives	Produces recommended alternatives including alternative futures based upon existing situation and resources to increase general understanding, situation understanding, and/or new capabilities.

Package Name	Package Description
21. Trans-frames	Rules of thumb supporting reactive behavior that provide the imagination of possible outcomes of a current situation or a small perturbation of the current situation.
22. Trans-frames Learning	Learn rules of thumb generated and generalized from alternative futures generation and evaluation processes.
23. Process Planning and Analysis	Analysis of a variety of processes in variety of domains, incorporates analogical and inductive reasoning to propose and evaluate new processes.
24. Process Learning	Incorporates taught or created new processes and learns application contexts and boundaries.

Foundation Unification Packages

Foundation packages supporting the *Unification* dimension of the INCOG framework are described in Table 3. In particular, the *Creativity Generator* and *Exploration and Discovery* packages provide key support to the *Creativity* level of capability in the framework. The *Situation Understanding* package supports multiple capabilities, especially the *Understanding* capability, which also draws on the learning packages (*Plan Learning* and *Skill Learning*). The *Behavior/Actions Context Generation* package supports all unification levels from *Understanding* to *Consciousness*. The *Planning* package is used at the *Plan* level and above, while the *Skills* package is accessed via all levels of unification, including the *Reactive*.

Table 3. Foundation Unification Packages

Package Name	Package Description
25. Creativity Generator	Self-motivated analogical and inductive reasoning to create, maintain, and obsolete understanding based upon semantic relations between knowledge including formal, system, conceptual, and mental models.
26. Exploration and Discovery	Explore and discover knowledge about internal and external worlds, enhance understanding and alternative behavior.

Package Name	Package Description
27. Situation Understanding	Support for situation understanding in any domain based upon a general model and process.
28. Behavior/Actions Context Generation	Generates semantic relations to link Understanding component of knowledge to possible behavior/actions in context of existing or hypothetical goals, objectives, and existing behavior/actions concept libraries.
29. Plan	Supports planning with generalized planners for a variety of domains.
30. Plan Learning	Plan learning and unlearning by experience, discourse and instruction, e.g., captures hypothetical and executed plans and builds performance models and semantic relationships to support plan knowledge creation, maintenance, and obsolescence.
31. Skills	Built-in and acquired skills application and maintenance.
32. Skill Learning	Skill learning and unlearning by experience, discourse, and instruction.

Foundation Global Mechanisms Packages

Certain cognitive packages provide functionality supporting most, if not all, aspects of the top-level architecture. Interface packages for working and long-term memory are the best example of such global foundation mechanisms since practically all functional components will need to access and/or store knowledge in memory. The other global mechanism packages (*Programming Model Observer*, *Self-Programming Module*, *Explanation of Behavior,* and *Inductive Repair*), while not as essential as memory access, can provide an invaluable service to many architectural components by enabling assessments of performance and correction of functional and knowledge deficiencies.

Table 4. Foundation Global Mechanisms Packages

Package Name	Package Description
33. Working Memory Interface	Access to distributed working memory representing known state of world, problems spaces, alternative futures, and distributed components. Working memory includes iconic and symbolic representations for states, goals, sub-goals, plans, skills, actions, alternative futures, events, actors, and objects of present interest.
34. Long Term Memory Interface	Access to long-term memory for any process supporting multi-level mind processing. Creates, modifies, and obsoletes knowledge structures based on instruction, experience, and creativity provided by other global mechanisms.
35. Self-Programming Module	Reasoners are supported by program and code generators and existing problem space libraries.
36. Programming Model Observer	Self-generated model observers specialized to each model type. Create, modify, and obsolete visibility to embedded Reasoners.
37. Explanation of Behavior	All packages support explanation of knowledge used, processes invoked, and status; this package integrates and presents explanations in the context of system model, world model and goals, objectives, plans, tasks, skills, and actions.
38. Inductive Repair	Inductive repair reviews system performance with system model and determines process obsolescence and redaction of working memory and knowledge base structures generated.

Foundation Reasoners Packages

The *Reasoning Mechanisms* dimension of the INCOG framework is supported by separate packages for each of the principal capabilities in the dimension, as described in Table 5.

Table 5. Foundation Reasoners Packages

Package Name	Package Description
39. Deductive	Deductive reasoning is a branch of cognitive psychology investigating cognitive systems ability to recognize a special relation between statements.
40. Nonmonotonic	The subject matter of nonmonotonic reasoning is that of developing reasoning systems that model the way in which commonsense is used by humans.
41. Probabilistic	Probabilistic reasoning is the formation of probability judgments and of subjective beliefs about the likelihood of outcomes and the frequencies of events.
42. Inductive	Inductive reasoning is reasoning from facts to a generalization about them. Inductive reasoning may infer simple empirical generalizations. Induction is one kind of inference that introduces uncertainty, hence conclusion must validated through modeling and simulation and/or experiment.
43. Abductive	Abductive reasoning is reasoning in which explanatory hypotheses are formed and evaluated.
44. Analogical	Analogy is (1) similarity in which the same relations hold between different domains or systems; (2) inference that if two things agree in certain respects then they probably agree in others.

Foundation Domain Packages

The *Reasoning Domains* dimension of the INCOG framework is supported by a group of packages whose mapping to the principal capabilities in the dimension is generally obvious, per the descriptions in Table 6. While spatial and temporal modeling are combined in a single package here due to their similarities, this package (45) can be expected to decompose with specialized components for those aspects of spatial and temporal reasoning that benefit from separate treatments. Advanced capabilities involving spatial, temporal, and causal modeling are delegated to a separate package (47) for scientific modeling. Lower-level support for spatial and temporal modeling of sensors and effectors also has separate support, from the *Kinesthetics and Perceptual* package (48). A package for commonsense assessment of situations (49) is supplemented with one focused on learning of commonsense in social and economic situations (50) to emphasize the special

needs of this capability. A final reasoning domain package (51) is allocated to resources, their modeling, monitoring, and utilization.

Table 6. Foundation Domain Packages

Package Name	Package Description
45. Spatial and Temporal Relations	Support for spatial and temporal reasoning, invocation, and working memory; and for knowledge creation, maintenance, and obsolescence.
46. Causality Relationships	Support for causal relationships in working memory and knowledge creation, maintenance, and obsolescence including current and past episodes.
47. Math, Physics, Chemistry, Biologic, and Other Sciences	Support for scientific concepts, analysis, modeling, and simulation for application to a variety of problems.
48. Kinesthetics and Perceptual	Support for embodiment in different applications of the artificial mind for adapting sensors and actuators of different types; also includes perceptual processors for 2D and 3D image processing including feature extraction.
49. Commonsense in Social and Economic Situations	Supports commonsense assessment of current situation; supports commonsense rules of thumb attributes, relationships, and behavior.
50. Commonsense in Social and Economic Situations Learning	Supports development of commonsense, social and economic knowledge from experience and instruction.
51. Resources	Models and evaluates external resource states, projects resource availability, evaluates known history and provides goals, objective plan, tasks, and actions to meet current objectives and possible future objectives.

Foundation Discourse Support Packages

Separate foundation packages are identified to support linguistic discourse, including language generation and understanding, as described in Table 7.

Table 7. Foundation Discourse Support Packages

Package Name	Package Description
52. Discourse Generator	Generates proposed discourse to support a specific context, goals and objectives. Includes processes to redact previous interpretations of linguistic inputs.
53. Context Interpreter	Identifies current context for discourse and provides appropriate semantics links to relevant knowledge, tracks threads of context associates with each goal and objective, and discourse episodes.
54. Listener and Parser	Semantically annotates voice or text episodes with linguistic and emotion expression relations.
55. Word, Phrase, Sentence	Provides word, phrase, sentence, and episode syntax and semantics in context.
56. Human Emotion Interpreter	Human and machine composite linguistic and visual interpreter for episode emotion, e.g. mood, emphasis, etc.
57. Emotion Expression	Identifies emotion expression expressed though voice or body state or movement, embedded emotion signals.
58. Human Emotion Library	Provides interpretation of human emotion expressed through speech, language, and visual perception and provides emotion expression learning from experience.
59. Machine Emotion Library	Provides voice, text and visual mechanisms to express current emotional state of the machine to support discourse and team activity.
60. Command, Instruction Interpretation, and Dialog	Interprets commands and instructions in context of current working memory model, maintains an external dialog as required.
61. Speech and Text Processing	Semantic interpretation of external speech and text inputs and internal stream of consciousness available to all processes with linguistic inputs.

Example Package Elaborations

A diagram of a possible functional decomposition of the Situation Understanding package (27) is provided to illustrate how a functional package may be decomposed into other functional components.

Figure 20 shows numerous aspects of one possible decomposition of situation understanding derived from a model of situation awareness used in information warfare [AGHS 2001]. The use of the function *Generate Alternative Futures* in this decomposition illustrates how distinct functional packages in the INCOG architecture may overlap, as this aspect of situation understanding would naturally call upon capabilities in the *Alternative Future Generator* package (19).

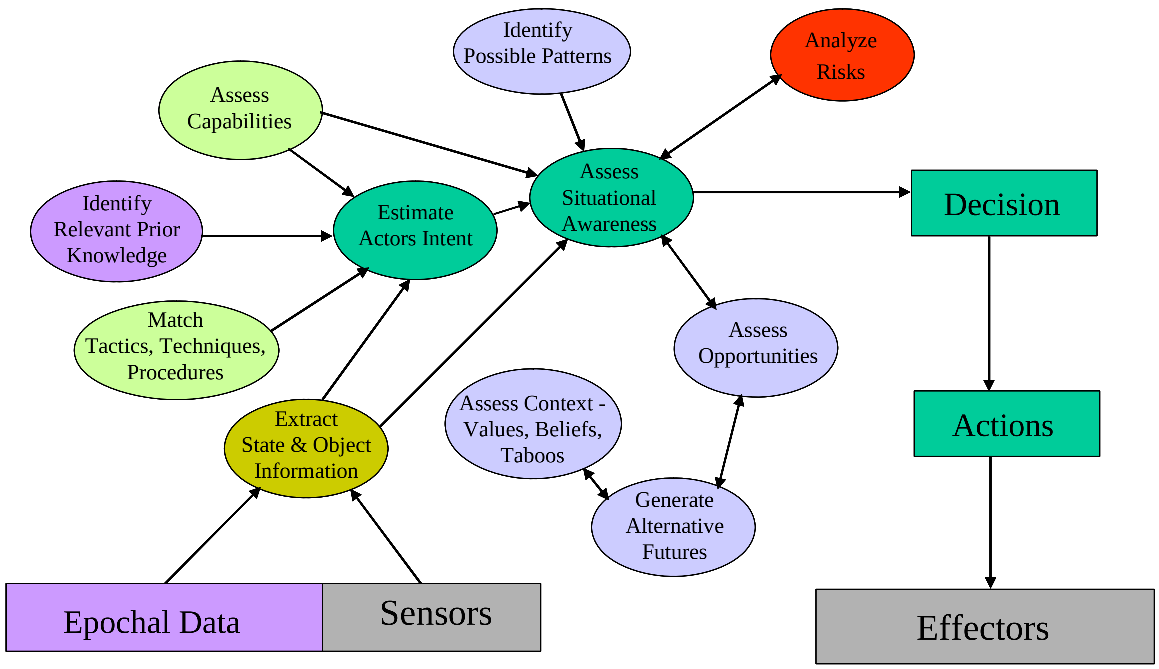

Figure 20. Situation Understanding Model (27)

A particular functional package description might also be elaborated via textual descriptions of the how the package functions in support of cognition. A brief textual elaboration of the *Self-Conscious Emotion Generator* package (12) is provided here to illustrate and to better explain its distinctive functionality.

Self-Conscious emotion addresses: To what degree are the events, actions, and object attributes in machines cognition window in compliance with learned values, censors, ideals, and taboos [ORCLCO 1988]. The six-level model of the human mind developed by Professor Marvin Minsky considers long-term memory associated with learned values, censors, ideals, and taboos as drivers of the self-conscious emotion level of the mind [MNSKY 2004]. There is a valence to the emotion and a resulting degree or strength met-

ric for preference [ORCLCO 1988]. In the context of discourse, selected use of vocabulary communicates valence and preference strength; this is true for human voice, text, and video inputs, and could be true for cognitive system outputs, voice, and multi-media. Dynamic inputs affecting self-conscious emotion include alerts and status notification from all other components of the cognitive system. The Self-Conscious Emotion provides motivation to change focus as required.

Cognitive Architecture Ontology

Any cognitive system that approaches human-level cognitive capacities will require a language (or languages) to express concepts, reason with them, and communicate them to others. And, the terms (words) of that language, which represent individual concepts, will need some representation of their semantics (or meaning) in order to enable meaningful processing of knowledge expressed in the language, as well as for communicating such meanings effectively to other cognitive agents, who may not share the same concepts or language.

An ontology is a formal specification of the semantics of terms used in a vocabulary. A popular definition by one of the pioneers in ontology development is:

> *In the context of knowledge sharing, I use the term ontology to mean a specification of conceptualization. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. This definition is consistent with the usage of ontology as set-of-concept-definitions, but more general. [GRUBER, 1993].*

Ontologies are useful as part of an architectural specification for cognitive agents in several respects. An ontology (or set of ontologies) can establish a uniform vocabulary for use by a community of cognitive agents. This can enable considerable precision in communications among the agents of such a community. But, just as important is that an ontology provides a specification of the semantics of a vocabulary to be used *within* a cognitive agent. In particular, in the context of a cognitive architecture conceived of as composed of many distinct functional components communicating with each other to realize higher-level cognitive processes, an ontology (or set thereof) is essential to establish common understanding of terminology used by the different components. Such ontologies not only enable communications, they enable a modular architecture in which different realizations of a particular type of component (or functional package) can be swapped in and out to explore alternative implementations without any need to adjust the other components that interact with it.

Complex architectures, such as the one sketched here, may define a single overarching ontology to cover all information exchange, or may define separate ontologies to cover distinct vocabularies used between subgroups of architectural components. Even a single ontology may be divided into parts based upon natural groupings of related concepts. Following is a list of a number of such concept groupings (here referred to as libraries) that have been identified as required to support certain components of the presented INCOG architecture.

- Abstract Concept Libraries
    - Sets, Sequences, Numbers and Arithmetic, Relations, Mathematics, . . .
- Human Natural Language Libraries
- Human Emotion Foundation Library
    - Human Emotional Expression Library
        - Vocal intonations, words, facial expression, and gesture
- Machine Emotion Foundation Library
- Multiple Domain Concepts Ontology Library
        - http://www.daml.org/ontologies/ , http://www-ksl-svc.stanford.edu:5915/
- Epochal Dialog Ontology Library
- Epochal Frame, Trans-frame, … Library
- Future Prediction Ontology Library by Domain
- Learning Strategies Ontology Library

Abstract concepts are grouped separately from concrete concepts, such as those that refer to physical objects, their properties, and relationships. Ontologies for human natural language are distinguished in order to allow artificial cognitive agent ontologies some freedom from natural language constraints, while still allowing them to communicate with humans using a variety of human natural languages. Emotion libraries may also define different variations on emotional expression for machines than they do for humans, as reflected in the dual emotion libraries listed. Separate ontological categories may also be distinguished for epochal dialogs and frames, as well as for types of predictions and learning strategies, as listed here. The *Multiple Domain Concepts* ontology library is a catch-all category that can capture ontological groupings for many other domains of knowledge, some of which can be found in special topic ontologies documented at the cited reference websites.

Learning Mechanisms

Learning has long been argued to be an essential capability in any system that aspires to intelligence, much less to human-level cognitive performance. There are many diverse learning strategies and different ways of categorizing them. The INCOG framework divides learning strategies into the following six categories, based roughly on the amount and types of inference required of the learner during the course of learning:

- Rote
- Stimulus Response

- Process Learning
- Instruction
- Discourse
- Exploration and Discovery

Rote learning requires practically no inference. Stimulus response learning inherently involves classification, i.e., of the presented stimuli to a generalization of them which enables a conditioned response to stimuli of that type. Process learning may include a wide variety of inferential capabilities in order to support learning from mistakes. Inference involved in exploration and discovery can range from the quite minimal, in learning simple facts about the environment, to the most demanding, in devising, executing, and evaluating experiments to discriminate among alternative scientific theories.

The higher-level learning strategies may involve many diverse learning mechanisms, such as the following:

- Analogical Reasoning
- Case-based Reasoning
- Classification and Regression Trees
- Genetic Algorithms
- Genetic Programming
- Neural Networks
- Rule Induction
- Statistical Pattern Recognition – Bayesian, Nearest Neighbor

Although descriptions of these mechanisms are beyond the scope of this paper, the reader can find such descriptions in general machine learning texts, such as [MITCHL 1997].

New Model Capabilities for Learning

The most capable integrated cognitive systems will have special abilities to manipulate their own learning mechanisms (or processes) and related knowledge representations. In particular, such learning systems will be able to generate new knowledge representations and related learning processes, e.g. rules or others, in various domains. But, that does not mean that such a system needs to begin as a "tabula rasa," free of learning processes and knowledge. Rather, initial learning processes and knowledge stores may be used to bootstrap learning in such systems. And, it is expected that even the most capable cognitive systems will require built-in knowledge and process generators to allow linguistic and other perceptual inputs to be dynamically related to its semantic needs. However, all learning processes and their knowledge representations will be evaluated over time, nominated, tested, validated, and discarded as warranted. In order to support these evaluations and revisions to learning knowledge and processes, such cognitive systems will require some type of self-programming model. They will need to be able to reprogram themselves to learn most effectively in different contexts.

Linguistic Learning

Language is given special attention in this strawman cognitive architecture because linguistic capabilities are seen as essential for some higher-level cognitive processes. We will not try to argue this view here, although it is certainly supported by our experience with natural cognitive systems. Linguistic capabilities are especially important for many types of learning, which are enabled by linguistic communication. In this section, we briefly review how learning from external information sources via language fits into the strawman cognitive architecture.

Basic processes for knowledge acquisition from external linguistic sources are illustrated by the diagram on the left side of Figure 21. This shows how different levels of linguistic representations may enter the language processing facilities at different levels. Speech input arguably requires special auditory processing (*Voice Recognition*) before it is recognized as speech. Recognized linguistic elements of speech (e.g., phonemes and words) can then be passed on to a *Natural Language Processing (NLP) Algorithm*. Simplified text input can bypass the voice processing algorithms, and may be input directly to the NLP algorithm in some cases. Machine encoded text input, in particular, requires no special sensory recognition capabilities in machines, although humans will require text recognition processing (typically visual) to prepare written text for NLP. More structured text, such as XML tagged text in the News Industry Text Format (NITF) might bypass basic NLP algorithms, although it may still require processing to establish “episode context,” e.g., organizing facts into ordered episodes. The most highly structured text, such as that used in intelligent software agent communication or knowledge bases might even bypass the *Episode Context Generator*. But, even at this level, linguistic information will require *Use Dependent Context Analysis* to establish its relevance to potential use in behavior. Hence, the diagram shows all input linguistic information going through this processing stage before it emerges as *use dependent* knowledge that can be effectively used in *Unified Behavior*.

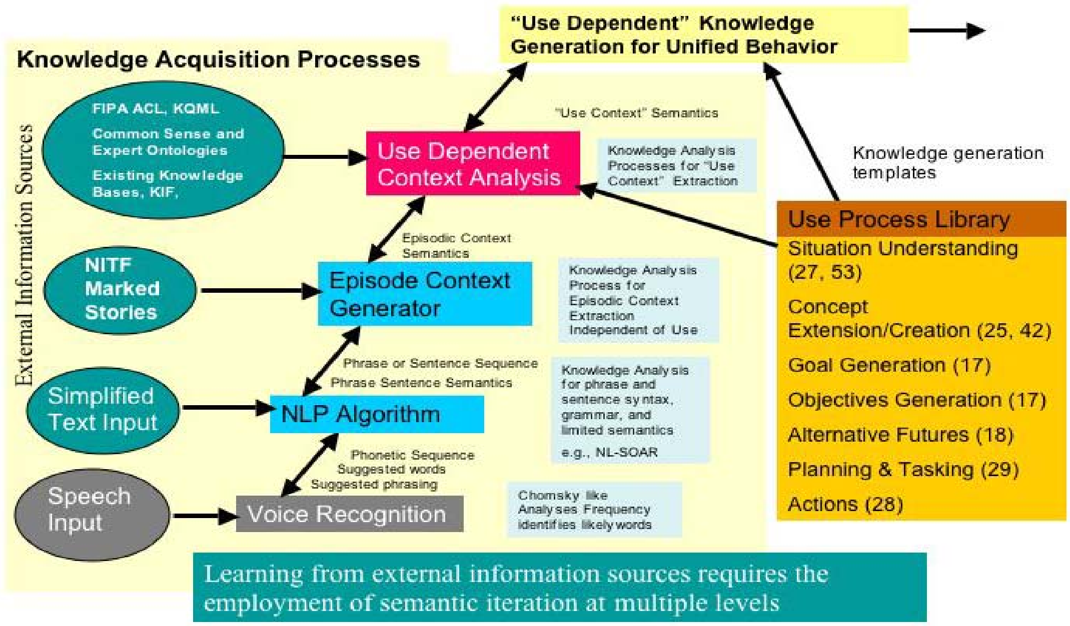

Figure 21. Relationship Between Linguistic Learning and Unified Behavior - Step 1

The right side of Figure 21 shows a *Use Process Library*, which provides functional capabilities for both *Use Dependent Context Analysis* and *Use Dependent Knowledge Generation* drawn from the functional packages of the strawman architecture. Library components are labeled here with the numbers of the utilized functional packages. Altogether, this figure illustrates the first main step in linguistic learning – the recognition and contextual understanding of linguistic inputs. The next step, illustrated by, shows how such linguistic knowledge is stored in the *Knowledge and Process Library*.

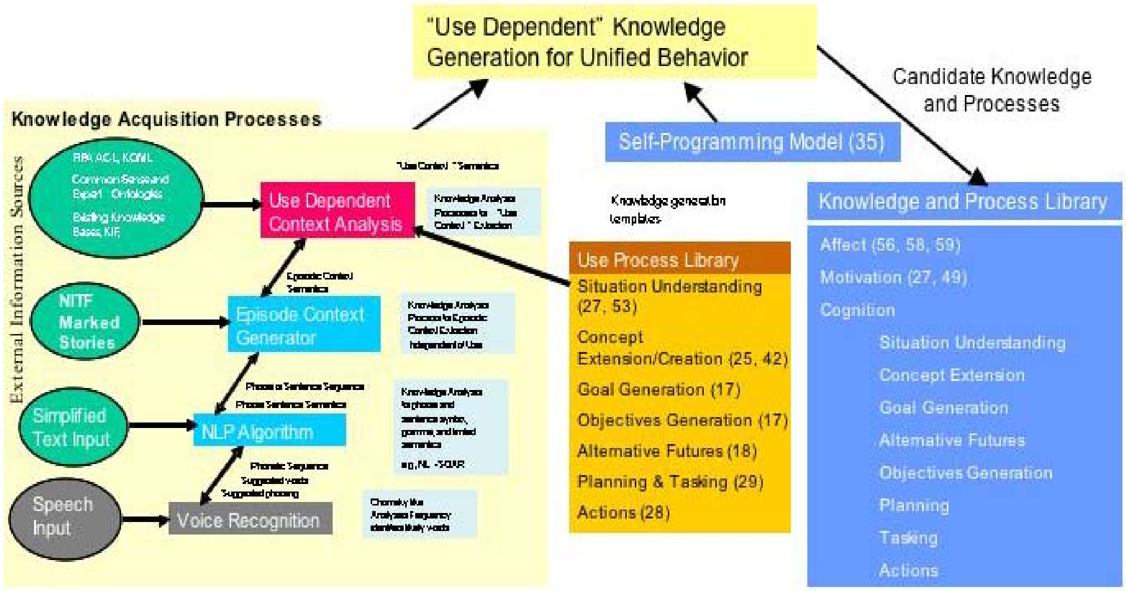

Figure 22. Relationship Between Linguistic Learning and Unified Behavior - Step 2

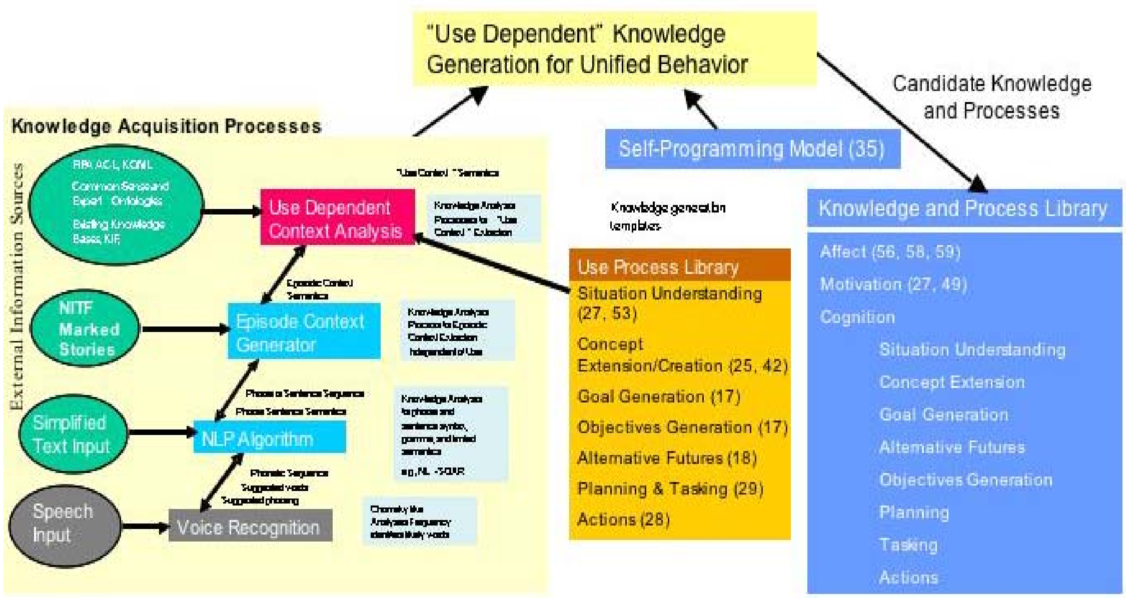

Figure 22 illustrates the use of a *Self-Programming Model* to generate process knowledge that is then stored in the *Knowledge Process Library* to alter or augment existing knowledge about how to do things, including knowledge about how to learn. This ability, not just to learn new facts, but to learn how to use knowledge and to learn new and revised

processes, cognitive and otherwise, is characteristic of the highest levels of learning capabilities.

Learning Expectations

Progress in learning systems will require systems capable of self-generation of knowledge and related processes. High-level cognitive learning requires abilities to both validate candidate knowledge and to unlearn putative knowledge and processes that are invalidated. Learning must be part of the integrated cognition architecture.

Today's "intelligent" or "expert" systems depend primarily on structured rules (possibly based on the formalized declarative semantics of an ontology), which are hand coded into knowledge and processes. Tomorrow's approach should be able to use a core of such hand coded knowledge to bootstrap the cognitive system, which can then acquire new information via structured text or simplified English inputs. Future, more advanced, cognitive systems will be able to use free form text available from libraries and the World Wide Web, as well as speech inputs and interactive dialogs to acquire knowledge. And, such advanced learning systems would, by no means, be limited to strictly language-based learning. Text and speech will be supplemented with pictures and diagrams. And, real world sensory inputs from visual, auditory, and other sensors will support other modes of knowledge acquisition and learning.

Structure Charts for Composite Cognitive Functions

The capabilities of the functional packages of the strawman cognitive architecture may be combined to create higher-level functionality, or may themselves be composed out of other functional packages. In this section, a couple of examples are provided of such composite cognitive functions. Structure charts are used to display the relationships between packages and the flow of information among them.

Apparent Conscious Behavior

Cognitive agents that exhibit apparently conscious behavior are likely to be drawing upon the functional cognitive packages shown in Figure 23. This structure chart shows inputs of alerts, and other status data, along with linguistic information annotated by emotional and other contextual information. These are accepted by the *Behavior Adjudication and Unification* package, which gets input on high-level goals and taboos from *The Builders Imprint*, and hands off its behavioral assessments to the *State Transition Preference Generator*. Preferences among competing goals are adjudicated in collaboration with *Process Planning and Analysis*, and the results propagated to the *Discourse Generator*, and others. The *Discourse Generator* is shown as making use of the *Machine Emotion Library* to assess and impress the appropriate emotional tone on discourse. Each of these processes will make use of *Working Memory* and *Long-Term Memory* as warranted.

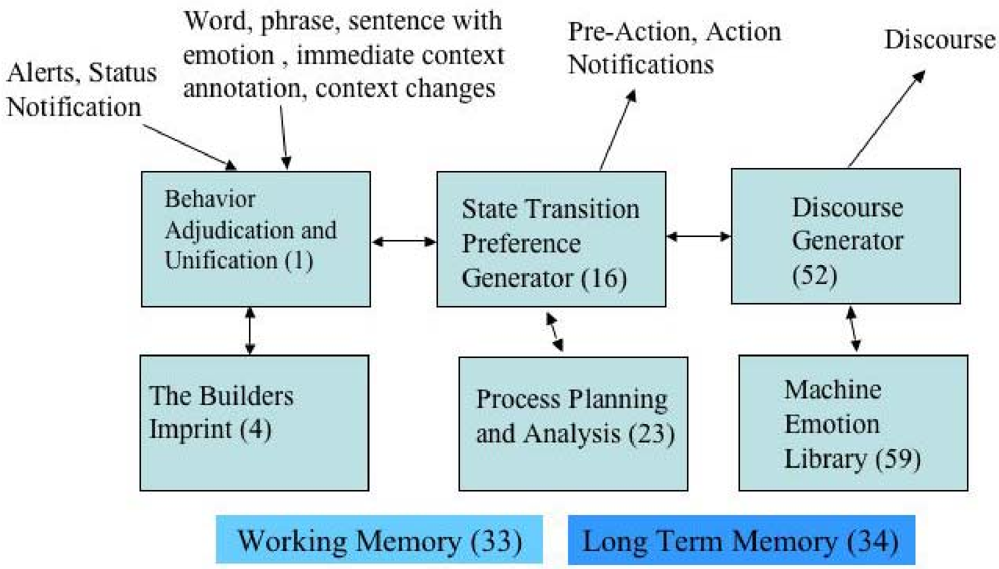

Figure 23. Static structure chart for apparent conscious behavior

Self-Reflective Package

An example of decomposition of a functional package into supporting functional packages is shown in Figure 24. Self-reflection is shown as being initiated by inputs from discourse, alerts, status notifications, actions, and planned actions, which are processed first by the *Cognitive Component Performance Evaluator*. In the context of self-reflection, the relevant inputs will likely concern perceptions/alerts on the self, including those generated by external actors. This uses the *Goals*, *Objectives*, *Metrics* package as a basis for evaluating self-performance, and shares its assessments with the *Alternative Future Generator*. This generates relevant possible alternative futures using the *Transframes* packages to determine possible outcomes of the current situation or small perturbations thereof. The resulting possible alternative futures relative to current situation, goals, and capabilities are processed by the *Recommended Alternatives* package to produce recommended alternatives. *Process Planning and Analysis* is used as warranted for analyzing relevant processes that may affect alternative futures of interest.

The result of this whole self-reflective process is a set of self-reflective alerts and/or status notifications, along with recommended actions when warranted. When self reflective performance assessments are neutral, they may not warrant any particular action or change of action. Negative self-assessments may require corrections to current behavior or adjustments to knowledge and process libraries to promote future corrections. Positive self-reflective assessments may serve to reinforce behavior. In any case, the output self-reflection results may trigger further analysis of the basis for the observed performance in order to avoid or repeat similar results in the future.

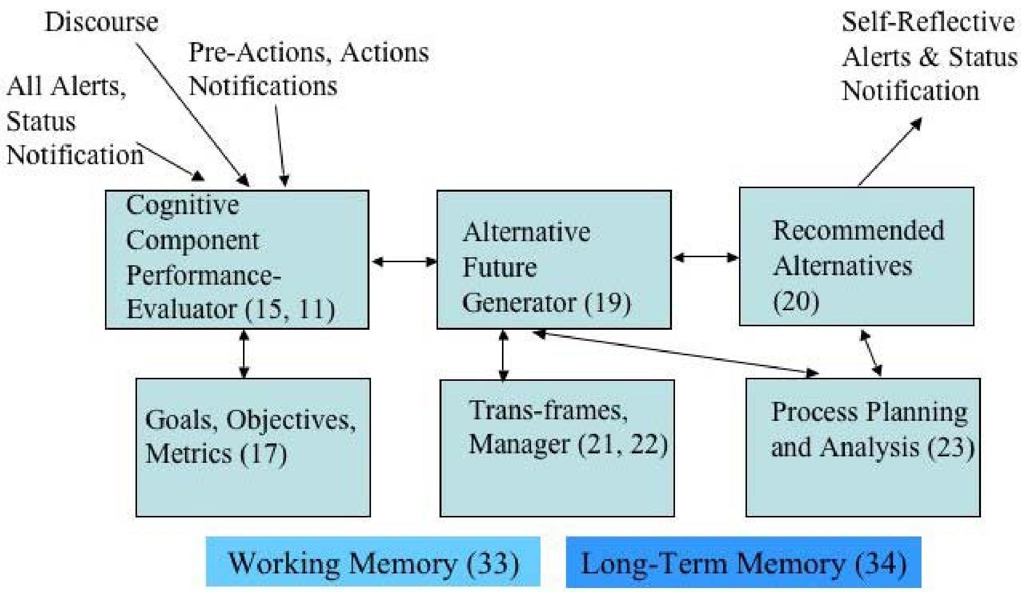

Figure 24. Static Structure Chart for Self-Reflective (6) Package

References

[AAAI 2004] AAAI "Machine Learning" website downloaded January 2004 from: http://www.aaai.org/AITopics/html/machine.html

[AGHS 2001] Alberts, D.S., Garstka, J.J., Hayes, R.E., and Signori, D.A. 2001. *Understanding Information Age Warfare*, DoD Command and Control Research Program (CCRP), Assistant Secretary of Defense (C3I). See URL: http://www.dodccrp.org

[ANDLE 1998] Anderson, J. and Lebiere, C. 1998. *The Atomic Components of Thought*. Mahwah, NJ: Erlbaum.

[BOBROW 1985] Bobrow D.G. (Ed.), 1985. *Qualitative Reasoning about Physical Systems*. Cambridge, MA: MIT Press.

[BRACH 2002] Brachman, R., "Developing Cognitive Systems. A Convergence of Thinking." Information Processing Technology Office (IPTO) Defense Advanced Research Projects Agency, briefing presented August 15, 2002.

[BRTMN 1987] Bratmann, M.E. 1987. *Intentions, Plans, and Practical Reason.* Cambridge, MA: Harvard University Press.

[EDEL 1987] Edelman, G. 1987. *Neural Darwinism: The Theory of Neuronal Group Selection*. Basic Books.

[FFG 1994] Forbus, K., Ferguson, W. Gentner, D. 1994. "Incremental Structure Mapping." 313-318. In *Proceedings of the 14th Cognitive Science Society*.

[GPSML 2003] Goertzel, B., Pennachin, C., Senna, A., Maia, T., and Lamacie, G. 2003. "Novamente: An Integrative Architecture for Artificial General Intelligence." In *Proceedings of the International Joint Conference on Artificial Intelligence-2003*.

[GRUBER 1993] Gruber, T.R. 1993. "A translation approach to portable ontologies." *Knowledge Acquisition, 5(2),* pp. 199-220.

[HOBMOR 1985] Hobbs, J.R. and Moore, R.C. (Eds.). 1985. *Formal Theories of the Commonsense World,* Norwood, NJ: Ablex Publishing Co.

[HUBER 1999] Huber, M. 1999. "JAM: A BDI-theoretic Mobile Agent Architecture," 236-243. In *Proceedings of the Third International Conference on Autonomous Agents and Multiagent Systems*.

[HUMHOL 2003] Hummel, J. and Holyoak, K. 2003. *Relational Reasoning in a Neurally-plausible Cognitive Architecture: An Overview of the LISA Project.* Los Angeles, CA: University of California.

[IKLW 1998] d'Inverno, M., Kinny, D., Luck, M. and Wooldridge, M. 1998. “A Formal Specification of dMARS.” In Singh, Rao, and Wooldridge (Eds.): *Intelligent Agents IV*, 155-176. New York: Springer Verlag.

[LSAS 2002] Langley, P., Shapiro, D., Aycinena, M., and Siliski, M. 2002. *A Value-Driven Architecture for Intelligent Behavior*. Stanford, CA: Stanford University.

[LHKD 1994] Lee, J., Huber, M., Kenny, P., and Durfee, E. 1994. “UM-PRS: An Implementation of the Procedural Reasoning System for Multirobot Applications.” In *Proceedings of the AIAA/NASA Conference on Intelligent Robotics in Field, Factory, Service, and Space: 509-519. Houston, TX*. American Institute of Aeronautics and Astronautics/National Aeronautics and Space Administration.

[MCM 1983] Michalski, R., Carbonell, J. and Mitchell, T. 1983. “An Overview of Machine Learning,” Chapter I in *Machine Learning*. Palo Alto CA: Tioga Publishing Company.

[MINKER 1993] Minker, J. 1993. “An Overview of Non-Monotonic Reasoning and Logic Programming,” *Journal of Logic Programming, Special Issue, 17*.

[MNSKY 2004] Minsky, M. 2004. *The Emotion Machine*, draft Downloaded January 2004 from URL: http://web.media.mit.edu/~minsky/

[MITCHL 1997] Mitchell, T. 1997. *Machine Learning*. McGraw-Hill Science/Engineering/Math.

[MUEL 1998] Mueller, E. 1998. *Natural Language Processing with ThoughtTreasure.* New York: Signiform.

[MUEL 1990] Mueller, E. 1990. *Daydreaming in Humans and Machines.* Bristal, England: Intellectbooks.

[NEWELL 1981] Newell, Allen. “The Knowledge Level,” *AI Magazine, Vol.2* (Summer 1981), pp. 1-20.

[NEWELL 1982] Newell, Allen. “The Knowledge Level,” *Artificial Intelligence, Vol 18* (1982), pp. 87-127.

[NEWELL 1993] Newell, Allen. “Reflections on the Knowledge Level,” *Artificial Intelligence, Vol 59, No. 1-2* (1993), pp. 31-38.

[ONR in prep] Ortony, A., Norman, D. A., and Revelle, W. (in preparation). “Effective Functioning: A Three-Level Model of Affect, Motivation, Cognition, and Behavior.” To appear in J.M. Fellous & M.A. Arbib,

Who Needs Emotions: The Man Meets the Machine. New York: Oxford University Press.

[ORCLCO 1988] Ortony, A., Clore, G.L., and Collins, A. 1988. *The Cognitive Structure of Emotions*. Cambridge. Cambridge University Press.

[ORTNY 2003] Ortony, A. 2003. “On Making Believable Emotional Agents Believable.” In R. Trappl, P. Petta & S. Payr (Eds.). *Emotions in Humans and Artifacts*. Cambridge, MA: MIT Press.

[PARBRU 2003] Parunak, V. and Brueckner, S. 2003. *Stigmergic Cognition in InCog*. Alexandria, VA: Altarum Institute.

[RLN 1993] Rosenbloom, P., Laird, J., and Newell, A. 1993. *The Soar Papers: Readings on Integrated Intelligence.* Cambridge, MA: MIT Press.

[RSE 1990] Reeke Jr., G.N, Sporns, O., and Edelman, G.E. 1990. “Synthetic Neural Modeling: The “Darwin” Series of Recognition Automata,” 1498-1530. In *Proceedings of the IEEE, Vol 78, No.9*.

[SHAST 1999] Shastri, L. 1999. “Advances In Shruti - A Neurally Motivated Model Of Relational Knowledge Representation and Rapid Inference Using Temporal Synchrony.” *Applied Intelligence 11(1)*, 79-108.

[SINGH 2003] Singh, Push. 2003. *The Panalogy Architecture for Commonsense Computing Brief Description*. Downloaded March 2004 from URL: http://web.media.mit.edu/~push/PanalogyBrief.pdf

[THGSHL 1997] Thagard, P. and Shelley, C. 1997. *Abductive reasoning: Logic, Visual thinking, and Coherence*, Philosophy Department, University of Waterloo, 1997

[WILKEL 2001] Wilson, R.A. and Keil, F.C. (Eds.). 2001. *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*. Bradford Books. URL: https://cognet.mit.edu/library/erefs/mitecs/

[WILSON 2004] Wilson, B. 2004. *The Machine Learning Dictionary for COMP 9414* Australia: University of New South Wales. downloaded January 2004 from: http://www.cse.unsw.edu.au/~billw/mldict.html

Abbreviations and Acronyms

AAAI	American Association for Artificial Intelligence
ACT-R	Adaptive Control of Thought or Atomic Components of Thought - Rational
AGHS	Advanced GPS Hybrid Simulator
AGIRI	Artificial General Intelligence Research Institute
AI	Artificial Intelligence
APOC	Activating-Processing-Observing-Components
BDI	Belief Desire Intention
C2	Command and Control
CCRP	Command and Control Research Program
CMU	Carnegie Mellon University
CSLI	Computer Science Learning Lab
DARPA	Defense Advanced Research Projects Agency
DARWIN	a core operating system
dMARS	Distributed Multi-Agent Reasoning System
DoD	Department of Defense
FCS	Future Combat System
FIPA ACL	Foundation for Intelligent Physical Agents Agent Communication Language
GPS	General Problem Solver
HRL	(formerly Hughes Research Laboratories, now just HRL Laboratories, LLC)
ICARUS	(not an acronym)
IDA	Institute for Defense Analyses

INCOG	Integrated Cognition
IPTO	Information Processing Technology Office
IRS	Intelligent Reasoning Systems
ISLE	Info Systems Learning
JAM	another version of PRS
KIF	Knowledge Interchange Format
KQML	Knowledge Query and Manipulation Language
LISA	Laser Interferometer Space Antenna
MAC/FAC	Many Are Called but Few Are Chosen
MIT	Massachusett's Institute of Technology
MITECS	MIT Encyclopedia of Cognitive Science
NITF	News Industry Text Format
NLP	Natural Language Processing
NL-SOAR	unified language capability built in the Soar cognitive architecture
NSI	Neurosciences Institute
ONR	Office of Naval Research
PRS	Procedural Reasoning System
SAGE	Self-Aware Adaptive Generalization Engine
SEQL	(apparently not an acronym)
SME	Structural Mapping Engine
SOAR	general cognitive architecture for developing systems that exhibit intelligent behavior
UC	University of California
UCB	University of California, Berkeley
UCLA	University of California, Los Angeles
UMPRS	University of Michigan implementation of PRS

URL	Uniform Resource Locator
US	United States
XML	Extensible Markup Language

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YY): October 2004

2. REPORT TYPE: Study

3. DATES COVERED (From – To):

4. TITLE AND SUBTITLE: Integrated Cognition – A Proposed Definition of Ingredients, A Survey of Systems and Example Architecture

5a. CONTRACT NUMBER: DASW01-98-C-0067/ DASW01-02-C-0012

5b. GRANT NUMBER:

5c. PROGRAM ELEMENT NUMBERS:

6. AUTHOR(S):
Robert M. Rolfe, Task Leader
Brian A. Haugh

5d. PROJECT NUMBER:

5e. TASK NUMBER: DA-5-2326

5f. WORK UNIT NUMBER:

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESSES:
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER: IDA Paper P-3855

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES):
DARPA
3701 N. Fairfax Drive
Arlington, VA 22203-1714

10. SPONSOR'S / MONITOR'S ACRONYM: DARPA

11. SPONSOR'S / MONITOR'S REPORT NUMBER(S):

12. DISTRIBUTION / AVAILABILITY STATEMENT: Approved for public release, unlimited distribution: 16 November 2005.

13. SUPPLEMENTARY NOTES:

14. ABSTRACT

Numerous cognitive scientists believe that a human-level thinking machine must be composed of potentially hundreds of distinct subsystems with different structures, reasoning and learning mechanisms, and knowledge representations —with these components and their inter-relationships defining a (or a family of) dynamic architecture(s). DARPA tasked IDA to help define a program, integrated cognition, to specify architectural strategies and develop infrastructure mechanisms (if needed) that enable computation based systems to achieve a reasonable facsimile of human cognition. IDA developed a new framework to understand the problem of integrated cognition, evaluated the current state of knowledge, and has proposed a straw man architecture to kick-start the effort. The benefits of integrated cognition for thinking machines would be immense. This is truly a DARPA hard problem, the solution of which will have huge potential payoffs, in particular, to realize intelligent C2, C4ISR, and robotic force elements.

15. SUBJECT TERMS

Cognitive systems, artificial intelligence, multi-level mind, machine knowledge, episodic memory, self-consciousness, self-reflective, reflective, deliberative, reactive, instinct, planning system, goal driven behavior, learning strategies, machine discourse, machine exploration, reasoning domains, reasoning mechanisms, integrated cognition.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER (Include Area Code)
Unclassified	Unclassified	Unclassified	Unlimited	90	Ms. Debra Amick / (571) 526-4163